NSS
Neutrosophic Sets and Systems, Vol. 90, 2025
University of New Mexico

Quantifying Indeterminacy in News: A Neutrosophic Method for Assessing Fact-Checking Outcomes

Dan Valeriu Voinea [1]*

[1] University of Craiova, Romania; dan.voinea@gmail.com

* Correspondence: dan.voinea@gmail.com

Abstract: Fact-checking news claims often encounters situations beyond simple true/false verdicts, involving partial truths, conflicting evidence, or insufficient information. Conventional binary or multi-class rating systems in both manual and automated fact-checking struggle to adequately represent this nuanced reality, particularly the degree and nature of indeterminacy. This paper proposes a novel framework grounded in neutrosophic logic to quantify truth, falsehood, and particularly indeterminacy in automated fact-checking outcomes. Neutrosophy explicitly models information using independent degrees of Truth (T), Indeterminacy (I), and Falsity (F), offering a richer representation than traditional logics. 'Indeterminacy' here encompasses ambiguity, vagueness, contradiction, uncertainty, and neutrality. Based on a systematic literature review spanning journalism, misinformation studies, and uncertainty modeling, we develop a theoretical model and outline a computational procedure for deriving (T, I, F) scores for news claims based on aggregated supporting, refuting, and unresolved evidence identified by automated systems. This approach allows claims to be assessed not just as true or false, but by how true, how false, and how indeterminate they are, given the available evidence, explicitly handling partial truths and contradictions. While primarily theoretical at this stage, the proposed framework offers a foundation for developing more transparent and nuanced AI fact-checking tools capable of communicating the inherent complexities and uncertainties in assessing real-world information, bridging neutrosophic theory with journalistic practice.

Keywords: Neutrosophic logic; Fact-checking; Misinformation; Uncertainty quantification; Automated journalism; Natural language processing;

1. Introduction

In an era of "fake news" and information disorder, accurately determining the veracity of news claims has become a pressing challenge. Professional fact-checking organizations have emerged worldwide to verify public statements, but their conclusions often defy simple binary classification. Many fact-checked claims exist on a spectrum between wholly true and wholly false [1]. Journalistic fact-checks frequently use nuanced ratings – such as *"mostly true"*, *"half true"*, or *"partly false"* – acknowledging that truth in news can be a matter of degree. For example, PolitiFact's six-level Truth-O-Meter ranges from *True* to *"Pants on Fire"* (extremely false), with intermediate labels signaling various mixes of accuracy and omission [2]. This gradation reflects a core insight from journalism studies: *not all claims lend themselves to a simple true-or-false verdict* [1]. Some statements contain elements of truth as well as falsity, or are unverifiable given available evidence. Misinformation research similarly recognizes "half-truths" – claims that are **partially true** but misleading or lacking context [3]. In fact, large-scale analyses of online news have explicitly categorized content as true,

false, or *mixed (partially true, partially false)* [4], underscoring that indeterminate truth values are commonplace in public discourse.

At the same time, there is a growing reliance on **automated fact-checking** systems to cope with the volume of dubious claims circulating online. Researchers in computer science and computational journalism have been developing algorithms to assist or mimic human fact-checkers [1]. These systems use techniques from natural language processing (NLP), information retrieval, and machine learning to identify claims and predict their veracity [5]. Notable efforts include claim spotting tools like ClaimBuster, which flags check-worthy statements [1], and claim verification models that compare claims against knowledge bases or evidence from the web. Datasets such as the **FEVER** benchmark (Fact Extraction and VERification) have framed fact-checking as a three-class problem – labeling claims as *Supported, Refuted,* or *Not Enough Info* based on evidence [6]. Similarly, real-world claims datasets like LIAR leverage multi-class truth scales from journalism (true, mostly true, half true, etc.) to train automated veracity predictors [2]. These efforts indicate an acknowledgment that automated fact-checkers must handle uncertainty or partial truth, at least by introducing a third category for *"not enough evidence"* or intermediate truthfulness. For instance, the FEVER dataset's **Not Enough Info** label explicitly represents cases where the system cannot determine truth or falsehood due to missing evidence [6]. More recently, the AVeriTeC dataset extended this idea by adding a *"Conflicting Evidence"* category for claims with **mixed supporting and refuting evidence** [7] – essentially flagging internal indeterminacy when sources disagree.

Despite these advances, most automated fact-checking pipelines ultimately reduce outputs to a single label or score, which may obscure the nuanced reality of the claim's status. A **binary or categorical label** (even a three-way label) simplifies the rich tapestry of information into a coarse verdict. Important questions remain open: *How confident is the system in a true/false judgment? If evidence is incomplete or contradictory, can we quantify that uncertainty? How can a machine express that a claim is "partly true" in a principled way?* These concerns align with longstanding calls in journalism to communicate not just facts but the **confidence or ambiguity** surrounding those facts [1]. Social science research suggests that fact-checks conveying nuance (e.g. "half true" conclusions) receive less public attention, as audiences gravitate to clear-cut answers [2]. Nevertheless, conveying uncertainty is critical for transparency and trust in both human and machine fact-checking. [8]

We propose a novel approach to represent and quantify the uncertainty in news fact-checking outcomes using the lens of **neutrosophic logic**. Neutrosophic logic, introduced by Smarandache , generalizes classical and fuzzy logic by explicitly accounting for *indeterminacy* alongside truth and falsehood [9]. Every proposition in neutrosophic logic is characterized by a triple **(T, I, F)** indicating its degree of truth, indeterminacy, and falsehood respectively in the real unit interval [0,1] (or more generally, in ⟨0,1⟩ allowing non-standard values) [9]. Latest research shows that the ranking order of these triplets, when determined by score, accuracy, and certainty functions, can change upon normalization [10]. Unlike a probability that must sum to 1 or a binary true/false assignment, a neutrosophic truth-value can simultaneously have non-zero truth and falsehood components, reflecting the possibility of **partial truth** and **contradiction**[9]. This makes it particularly well-suited to model fact-checking scenarios where evidence may be mixed or incomplete. By developing a neutrosophic framework for automated fact-checking, we aim to enable machines to output *"true to degree x, false to degree y, indeterminate to degree z"* for a given claim – a much richer outcome than current systems' binary or ternary decisions. Neutrosophy originates from philosophy, specifically studying neutralities and their interactions, challenging classical binary thinking by incorporating indeterminacy alongside truth and falsity. Crucially, unlike fuzzy logic or intuitionistic fuzzy logic where components are often dependent, neutrosophy allows T, I, and F to vary independently (e.g., in Single-Valued Neutrosophic Sets, the constraint is $0 \leq T + I + F \leq 3$), enabling the modeling of contradictory or incomplete information where, for instance, both T and F might be high.

2. Literature Review

2.1. Fact-Checking in Journalism and the Problem of Partial Truths

Fact-checking as a journalistic practice has expanded significantly in the past decade, driven by concerns over political deception and viral misinformation. Unlike traditional reporting, fact-checking involves evaluating specific factual claims and rendering a verdict on their accuracy. Research in journalism studies shows that fact-checkers rarely deal in absolutes of *"completely true"* or *"completely false"*. Instead, nuanced judgments are common, reflecting the complexities of context, wording, and evidence [1]. Graves observes that fact-checking organizations have developed rating systems to express intermediate truth values, such as "mostly true" for claims that are accurate but missing context, or "half true" for claims that mix truth and falsehood [2]. These systems acknowledge what academics and audiences alike know intuitively: reality is often gray. Many statements contain a kernel of truth buried in exaggeration or misinformation. A study by Tandoc et al. [11] notes that the term *"fake news"* itself is problematic because false information can be interwoven with facts, making it partly credible. In line with this, recent scholarship refers to **misleading content** that is *partially true* or *partially false*, rather than entirely fabricated [3]. For instance, a viral rumor might accurately cite a statistic but misattribute its cause – the overall impression given can be false even though some facts are correct.

The concept of **half-truths** has been explicitly highlighted in misinformation research. Singamsetty and Bhattacharyya define *half-truths* as statements that are *"partially true or manipulated to misrepresent the truth"* [3]. Such content is considered especially pernicious because it can be more believable and harder to debunk than outright falsehoods. Allcott and Gentzkow [12] similarly warn that pieces of false news often contain some true information, which can lend them unwarranted credibility. A Science study by Vosoughi et al. classified online news stories as *true*, *false*, or *mixed (partially true, partially false)* when analyzing their diffusion on Twitter [4]. Notably, they found false (and mixed) stories spread farther and faster than true ones, underscoring the viral appeal of sensational half-truths.

From the journalism perspective, a key challenge is conveying these nuances to the public. Audiences may prefer a simple true/false dichotomy, and "partly true" verdicts can be confusing or less satisfying. There is evidence that fact-checks with clear-cut rulings (*True*/*False*) get more engagement than those with intermediate rulings [2]. Yet, providing a truthful account often requires embracing complexity. A claim might be **accurate in one aspect but inaccurate in another**; professional fact-checkers address this by writing lengthy explanations in addition to issuing a rating. The fine-grained analysis is essential for transparency, but the bottom-line message is harder to communicate succinctly. This tension suggests that new methods are needed to express graded truth values in a way that is both rigorous and understandable.

2.2. Automated Fact-Checking: Capabilities and Limitations

As the volume of online content surged, researchers began exploring automation to assist fact-checkers and even perform verification in real time. **Automated fact-checking (AFC)** encompasses a pipeline of sub-tasks: detecting check-worthy claims, retrieving relevant evidence (e.g., documents or database entries), and assessing claim veracity by comparing the claim to the evidence [1,5]. Early efforts like the ClaimBuster system [13] focused on the first step – using machine learning to identify sentences likely to contain factual claims worth checking [1]. Subsequent work tackled evidence retrieval and claim verification, often casting the problem as a classification task (true or false) or as a natural language inference problem (determining if evidence *supports* or *refutes* a claim). A comprehensive survey by Guo, Schlichtkrull, and Vlachos notes that approaches range from using knowledge graphs and databases (for instance, verifying statistical claims against official data) to using textual entailment models that read web articles to see if they confirm or contradict the claim [5].

In practice, most automated fact-checking systems simplify the output to a **single confidence score or label**. For example, an AI model might output a probability between 0 and 1 indicating how likely a claim is true, which is then thresholded to produce a binary true/false prediction. More advanced systems, influenced by datasets like FEVER, adopt a *three-way classification*: *Supported, Refuted,* or *Not Enough Information* [6]. The FEVER dataset's inclusion of a "Not Enough Info" category

was a recognition that sometimes the correct verdict is *insufficient data*, rather than forcing a mistaken true/false decision. Similarly, the recent AVeriTeC dataset added a *"Conflicting Evidence"* label to capture situations where the collected evidence both supports and contradicts the claim [7]. This is essentially an admission of indeterminacy: the truth value cannot be resolved because the source information is inconsistent.

Despite these enhancements, automated systems still face significant challenges in handling uncertainty and partial truth:

- **Context and Nuance:** Automated checkers struggle with claims that require understanding context or subtleties that go beyond literal content [1]. A claim might be technically true but misleading without context – humans recognize this nuance and might rate it "Half True," but a straightforward AI might flag it as simply true. Current NLP models are not reliably able to make the kind of contextual judgment that a human fact-checker would, such as noting that a statistic is correct *but* used out of context to create a false impression [1]. This often leads to a gap where the system's output (e.g., "true") doesn't fully align with the nuanced verdict a human would give ("partially true, missing context").
- **Calibration of Confidence:** Machine learning classifiers output probabilities or confidence scores, but communicating that confidence is non-trivial. A system might be 60% confident that a claim is true – is that sufficient to call it true? Some research suggests incorporating a rejection option, where the model abstains if unsure. This relates to the idea of **calibrated uncertainty** in AI predictions. However, few fact-checking prototypes explicitly expose an "uncertainty" measure to end-users. Typically, the system either makes a prediction or not. The "Not Enough Info" category in FEVER can be seen as an attempt to model uncertainty by a categorical outcome (the system essentially says, "I can't verify this"). Yet, in reality this is a binary decision to abstain, rather than a graded expression of uncertainty.
- **Contradictory Evidence:** When an automated evidence retrieval pulls in multiple sources, it is possible some sources support the claim while others refute it. How should the system handle this? One approach is to aggregate all evidence and let the classification model resolve it – often, the model will lean one way or the other, effectively averaging out the conflict. But a more transparent approach might be to acknowledge the conflict explicitly. The aforementioned AVeriTeC dataset suggests the need to label such cases separately [7]. A human fact-checker in this scenario might conclude the claim is *"controversial"* or *"uncertain"* because experts disagree or data are conflicting. Most AI models, however, lack a mechanism to output "conflicting evidence" as a state. They would likely either output an intermediate probability (around 0.5, reflecting confusion) or randomly lean toward one side if forced to choose a class. Neither is satisfactory for clear communication.
- **Generalization and Novelty:** Many automated fact-checking methods rely on previously seen claims or evidence. For instance, checking against a database of known false claims works only if the claim (or close variant) has been seen and stored before [1]. Truly novel claims require live evidence gathering and reasoning, which is much harder and often results in higher uncertainty. Automation is currently better at detecting *known falsehoods* than adjudicating new, complex claims. This means that for novel claims, an AI system is inherently uncertain – it may find some related information, but often not a definitive refutation or confirmation. How to quantify and present that uncertainty remains a challenge.

In summary, while automated fact-checkers have made strides – leveraging NLP to verify simple factual claims with structured data, and even using transformer-based language models to analyze text – they still struggle with the *indeterminacy* that is intrinsic to many real-world verification tasks [1]. Current systems either oversimplify by giving a binary verdict (with hidden confidence), or use a coarse extra category like "Not verified." There is a clear gap in representing the nuanced continuum of truth values that human fact-checkers communicate in prose. This gap motivates exploring more expressive logical frameworks.

2.3. Modeling Uncertainty: Beyond Binary Logic in Verification

The need to represent uncertainty and partial truth in automated reasoning is not unique to fact-checking; it has been long recognized in fields like artificial intelligence, decision science, and knowledge representation. Over the years, several **multi-valued logic and uncertainty modeling frameworks** have been developed:

- **Fuzzy Logic:** Introduced by Zadeh [14], fuzzy logic allows truth values to range between 0 and 1, representing degrees of truth. In fuzzy set theory, an element's membership in a set (e.g., the set of "true statements") can be partial. Classic fuzzy logic, however, typically assumes a single degree that implicitly covers both truth and falsity (if truth = 0.7, false can be inferred as 0.3 in a two-valued fuzzy interpretation) [9]. It does not explicitly encode ignorance; a membership of 0.7 means 30% not a member (false) and no additional category for "undecided." Thus, fuzzy logic introduced gradation but still essentially deals with one continuous spectrum between fully false and fully true.
- **Intuitionistic Fuzzy Sets:** Atanassov [15] expanded fuzzy sets to include a degree of hesitancy or indeterminacy. An *intuitionistic fuzzy set* is characterized by a truth membership degree t and a falsity membership degree f, with the condition that $t + f \leq 1$. The difference $1 - (t + f)$ is then interpreted as the *indeterminacy* or *hesitation* margin [9]. This framework acknowledges that for some elements, one might not be able to fully specify membership or non-membership – the remainder up to 1 is an uncertain part. In essence, intuitionistic fuzzy logic has three components (T, F, and an implied I), but they are constrained by $T + F + I = 1$. If applied to claim verification, an intuitionistic fuzzy verdict might say: truth=0.6, false=0.2, leaving 0.2 as uncertainty (*lack of knowledge or undecided*). This is a step closer to what we need, though the fixed sum constraint can be limiting in cases of contradictory evidence (where one might feel both truth and falsehood have substantial support).
- **Probabilistic Reasoning & Belief Functions:** Bayesian probability theory would model uncertainty by assigning a probability p that a claim is true (and $1 - p$ false). But a single probability cannot easily express "*unknown*" separate from an even split of belief. If $p = 0.5$, is it because we truly have no idea (max uncertainty), or because evidence is evenly balanced? Classical probability doesn't distinguish the two scenarios – both yield 0.5. To address this, Dempster–Shafer **evidence theory** [16] allows assigning belief mass to *unknown*. In Dempster-Shafer theory, one can assign a mass *m_true*, *m_false*, and *m_uncertain* such that these sum to 1. It's akin to saying: "*I believe this much in true, this much in false, and this much I leave uncommitted*." The *uncommitted* (uncertain) mass can then be allocated to either truth or falsehood in different ways (yielding bounds like belief and plausibility). This framework has seen use in data fusion and risk analysis, and it aligns conceptually with our needs in fact-checking – we could assign, for example, 70% belief to the claim being true, 20% to it being false, and leave 10% as unresolved uncertainty. However, Dempster-Shafer theory, while powerful, has complex combination rules and can behave counter-intuitively in the presence of conflicts (the combination rule amplifies conflict into lower normalized support). It also does not explicitly allow *contradiction* to be represented other than by a roughly equal split of belief between true and false.
- **Three-Way Decision Theory:** A more recent development in decision sciences is the notion of three-way decisions or tri-partition of outcomes, entered by Yao [17]. This approach echoes the intuition of *accept*, *reject*, or *defer*. In classification terms, it means an classifier can output *Yes*, *No*, or *Unsure*. The third option allows one to defer a decision when confidence is low or information is insufficient [18]. Three-way decision frameworks have been applied to, e.g., medical diagnosis and information retrieval, where taking an uncertain action might be costly so an *"undecided"* option is preferable. In fact-checking, one can see the parallel: *support the claim, refute the claim,* or *refuse to decide (not enough info)*. The FEVER label set {Supported, Refuted, Not Enough Info} is essentially a three-way decision model in practice [6]. What three-way decision theory adds is a formal decision-theoretic or rough-set-based foundation

for when to choose the third option, often based on thresholds of confidence or risk minimization. Nonetheless, one limitation is that it remains a **categorical** decision – it doesn't provide a nuanced gradation within the categories. A claim labeled "Unsure" in a three-way model doesn't tell *why* or *how much* evidence was lacking, just that the system chose not to decide.

Each of the above frameworks contributes towards representing uncertainty in a structured way. However, **neutrosophic logic** goes a step further by combining *gradualism* (like fuzzy logic) with *explicit uncertainty representation* (like Dempster-Shafer or intuitionistic fuzzy logic) and even *tolerance for contradiction*. Florentin Smarandache's neutrosophic theory emerged in the early 2000s as a **generalization of many multi-valued logics** [19,20]. A neutrosophic proposition has a truth membership T, falsity membership F, and indeterminacy membership I, conceived as independent parameters in [0,1] (or even outside [0,1] in some formulations) [9]. The only constraint in a *single-valued neutrosophic set* (the most common practical version) is $0 \leq T, I, F \leq 1$, but notably $\boldsymbol{T + I + F \text{ is not required to equal } 1}$. This means a statement could be, for example, $T = 0.8, F = 0.7, I = 0.0$ – an admittedly odd case in classical terms, since $T + F = 1.5 > 1$, but neutrosophically this would indicate that from certain perspectives the statement is largely true, from others largely false, and overall there is little indeterminacy (the high overlap between truth and false components signals **contradiction**). In other words, neutrosophy can handle the **simultaneous partial truth and partial falsehood** in a way that intuitionistic fuzzy sets (which would force $T + F \leq 1$) would not allow. At the same time, if something is completely unknown, one can set I high and T, F near 0. The flexibility of this representation makes neutrosophic logic a compelling choice for modeling fact-check outcomes, which may indeed involve contradiction (different sources or experts disagreeing) and varying degrees of unknown.

2.4. Neutrosophic Theory and Applications

Neutrosophic logic is part of the broader neutrosophic set theory and neutrosophic probability theory developed by Smarandache and colleagues [9,21,22]. In a neutrosophic probability, one would say an event's chance of happening is T%, not happening F%, and an indeterminate chance I% (perhaps due to unknown factors). Neutrosophic sets and probability generalize the classic notions by introducing this indeterminacy component. The primary motivation is to better model real-world situations where incomplete and inconsistent information is present. Indeed, *"neutrosophic sets have been developed to represent uncertain, imprecise, incomplete, and inconsistent information existing in the real world"* [19,23]. This resonates strongly with the information state of many news claims being checked, where data may be incomplete (leading to indeterminacy) or sources conflict (leading to inconsistency).

Over the last decade, neutrosophic approaches have found applications in various domains. For example, in **decision-making problems**, interval neutrosophic sets have been used to improve multi-criteria decision analysis by capturing the decision maker's indeterminacy about criteria. [19] In **image processing**, "neutrosophic filters" have been designed to handle uncertain pixels (like noise) by classifying pixels into true-feature, indeterminate (noise), or false-feature categories and then cleaning accordingly. In **text and sentiment analysis**, researchers have begun to apply neutrosophic logic to handle ambiguity in opinions. Essameldin et al. proposed a neutrosophic inference system for sentiment analysis of social media, where each opinion's polarity was expressed with truth, indeterminate, and false components to reflect ambiguity in sentiment expressions [24]. By converting "crisp" inputs (e.g., a sentiment score) into a neutrosophic form and then applying inference rules, they were able to better classify emotions that were not purely positive or negative. This illustrates the practical benefit of neutrosophic logic in an NLP context: handling mixed sentiment (both positive and negative) parallels handling mixed veracity (both truthful and deceptive elements). For example, in sentiment analysis, researchers have successfully mapped outputs from standard tools (like VADER) into neutrosophic (T,I,F) scores, explicitly modeling neutral or ambiguous opinions. Refined Neutrosophic Sets have even been used to distinguish nuances like

'indeterminate positive' versus 'indeterminate negative' sentiment, suggesting potential for similar fine-grained analysis in fact-checking indeterminacy.

Another relevant application is in **knowledge-based systems** and **expert reasoning**. Neutrosophic logic has been used in medical diagnosis systems where symptoms might support multiple possible conditions (truth of one diagnosis does not entirely negate others – there is overlap and uncertainty). Researchers have found that neutrosophic representations can improve the interpretability of such systems by explicitly showing the indeterminacy part of the conclusion [24]. In one case, a rule-based classifier was enhanced with neutrosophic logic to deal with overlapping class boundaries, by assigning an indeterminacy value when a data point lay in the overlap region of two classes. This prevented forced but unreliable classifications and allowed a human decision-maker to see that the case was ambiguous.

Of particular note for our purposes is the concept of **refined neutrosophic sets**. Patrascu introduced a refined neutrosophic information model that splits the indeterminacy I into sub-components such as *ignorance*, *contradiction*, and *hesitation* [25]. *Ignorance* refers to a lack of information (e.g., no evidence available), *contradiction* refers to the presence of conflicting information (e.g., evidence both for and against), and *hesitation* might refer to difficulty in making a judgment even with information (perhaps due to fuzziness or equal balance). By refining I, one can pinpoint the nature of the indeterminacy. This is highly relevant to news fact-checking – if a claim is unverified, is it because nobody studied it (*ignorance*), or because studies found opposite results (*contradiction*)? Though our proposed framework in this paper does not delve deeply into refined categories, we acknowledge it as a possible extension (see Discussion). The mere existence of such work underlines that neutrosophic theory is being tailored to complex real-world reasoning scenarios where "*uncertainty*" is multi-faceted.

In summary, neutrosophic logic provides a rich vocabulary for describing the state of knowledge about a proposition. It has been applied in fields requiring careful handling of uncertainty and inconsistency, albeit it remains a niche approach compared to mainstream probabilistic or fuzzy methods. Importantly, we did not find prior literature directly applying neutrosophic logic to the fact-checking of news or misinformation – indicating that our attempt is novel. There is, however, a conceptual synergy between **neutrosophy and journalism**: both recognize the importance of the middle ground between true and false. Journalism has terms like "unverified," "disputed," or "partly true," while neutrosophy offers a systematic way to quantify those states. Building on the literature reviewed, we now move to construct a theoretical framework that marries these ideas, setting the stage for a neutrosophic model of fact-checking.

3. Theoretical Framework

Bringing together insights from the literature, our theoretical framework centers on the idea that each factual claim in news can be assigned a **neutrosophic truth value** $(\boldsymbol{T}, \boldsymbol{I}, \boldsymbol{F})$ reflecting the state of evidence regarding that claim's validity. We conceptualizes an automated fact-checking system that processes a claim, gathers evidence, evaluates support and refutation, and then produces a triple score. The components of this score are:

- **T (Truth Degree):** The degree to which the claim is supported by evidence and can be considered true (or accurate). A higher T means strong corroborating evidence exists for the claim's content.
- **F (Falsehood Degree):** The degree to which the claim is contradicted by evidence and can be considered false (or inaccurate). A higher F means there is strong evidence disproving or refuting the claim.
- **I (Indeterminacy Degree):** The degree of uncertainty, ambiguity, or indeterminacy regarding the claim's veracity. A higher I indicates that significant aspects of the claim remain unverified, the evidence is conflicting or incomplete, or the truth value cannot be definitively resolved.

The indeterminacy component (I) in neutrosophy is broadly defined, encompassing not just unknown information but also ambiguity, vagueness, neutrality, contradiction, paradox, or simply a

middle ground, depending on the context. This tripartite representation directly addresses the shortcomings identified in existing systems. Instead of outputting a single label like "Unsupported" or a probability like 0.6, the system might output, for example: **Claim** $\boldsymbol{X} = (\boldsymbol{T} = \mathbf{0.6}, \boldsymbol{I} = \mathbf{0.3}, \boldsymbol{F} = \mathbf{0.1})$, meaning the claim is mostly true (60% true, 10% false) with some residual uncertainty (30% not determined). Such an output can be mapped to a natural-language statement for end users, like "*mostly true with some open questions*." In another scenario, a claim might get$(\boldsymbol{T} = \mathbf{0.4}, \boldsymbol{I} = \mathbf{0.4}, \boldsymbol{F} = \mathbf{0.4})$ if evidence is evenly mixed for and against – indicating a contentious, unresolved claim. (Notably $T + I + F$ can sum to >1 here, which is allowed; this highlights contradiction since both T and F are relatively high). If a claim has not been studied at all, we might see $(\boldsymbol{T} = \mathbf{0.0}, \boldsymbol{I} = \mathbf{1.0}, \boldsymbol{F} = \mathbf{0.0})$– essentially "unproven" due to ignorance. These examples illustrate how the neutrosophic values provide a richer semantic than a single scalar or category.

Under the hood, how do we arrive at these (T, I, F) values? The framework assumes an underlying **evidence assessment process** typical of automated fact-checking: the system collects a set of evidence items $\mathrm{E} = \mathrm{e}_1, \mathrm{e}_2, \ldots, \mathrm{e}_\mathrm{n}$ related to the claim. Each evidence item could be, for instance, a news article, a database entry, an expert quote, or a prior fact-check, accompanied by some measure of reliability or relevance. We then need to evaluate *stance* – does each e_i support the claim, refute the claim, or is it neutral/not applicable? Many NLP models exist for stance detection or textual entailment that can perform this categorization, often with a confidence score. For our theoretical construction, we denote:

- S = the aggregate *support* for the claim from evidence.
- R = the aggregate *refutation* for the claim from evidence.
- U = the aggregate *uncertainty* or *unclear portion* of evidence.

These aggregates could be simple counts (e.g., number of sources supporting vs refuting), or weighted sums (taking into account source credibility or relevance), or even probabilities from a model (e.g., a model might give a probability distribution over "supports/refutes/neutral"). For simplicity, one can think of S, R, and U as follows: out of all relevant evidence pieces the system considered, some fraction came out supporting (that fraction contributes to S), some refuting (contributing to R), and the rest did not contribute a clear verdict either way (contributing to U). By definition, we can set a normalization $S + R + U = 1$ (i.e., distribute 100% of the evidence weight among support, refute, and inconclusive/neutral). This U captures ignorance in the *evidence space* – perhaps no source was found (then U is large), or sources found didn't directly address the claim.

Now, we map these evidence aggregates to neutrosophic components. A straightforward mapping is:

- **Truth degree (T)** = S (the proportion of evidence supporting the claim).
- **Falsehood degree (F)** = R (the proportion of evidence refuting the claim).
- **Indeterminacy degree (I)** = U (the proportion of evidence that is neutral or insufficient).

Using the earlier example: if out of 10 pieces of evidence, 6 supported, 1 refuted, and 3 were irrelevant or mixed, then $S = 0.6, R = 0.1, U = 0.3$. We would assign $(T = 0.6, F = 0.1, I = 0.3)$. This would correspond to a "*mostly true*" outcome. On the other hand, if 4 supported and 4 refuted and 2 were unclear, we get $S = 0.4, R = 0.4, U = 0.2, i.e., (T = 0.4, F = 0.4, I = 0.2)$ – an indeterminate case with strong contradiction (we might label it "unclear / debated"). If 0 supported, 0 refuted, and 10 were neutral/not found, then $S = R = 0, U = 1.0$: $(T = 0, F = 0, I = 1)$ meaning "unverifiable (no evidence)".

This initial scheme aligns with an *intuitionistic fuzzy* interpretation $(T + F + I = 1)$. However, neutrosophic logic allows more freedom which can be useful in certain scenarios. For instance, consider contradictory evidence from two equally authoritative sources: one says the claim is true, another says it is false. Arguably, we have *full support and full refutation*. In our evidence aggregation, we might set $S = 0.5, R = 0.5$ if we had two pieces (one support, one refute). That yields $T = 0.5, F = 0.5, I = 0.0$. But one might argue the indeterminacy is actually very high in such a case because we don't know which source is right. A different approach could assign overlapping truth and falsehood rather than splitting the difference. Neutrosophically, we could let $S = 1.0$ and $R = 1.0$ to represent "100% supported *and* 100% refuted" (two conflicting certainties), which then implies a contradiction.

Of course, if we feed those into our normalized scheme we cannot have S and R both be 1 under the $S + R + U = 1$ constraint. But neutrosophic logic suggests that perhaps the sum need not be 1 in cases of conflict – instead, we record the conflict by allowing T and F to both be high. One way to implement this is to maintain the normalized S, R, U for evidence coverage but separately record a *conflict measure* C. For example, C could be the degree to which evidence is bimodal – if both support and refute evidence exist above some threshold, C captures that. Then we might define neutrosophic values as: $T = S, F = R, and\ I = U + C$. Here C (conflict) is a part of indeterminacy, because contradiction is a form of indeterminacy (we don't know which side is correct). In the earlier case of one support/one refute, $S = R = 0.5, U = 0$, and say we set $C = 0.5$ (because the evidence is evenly split, conflict is high). Then $(T, I, F) = (0.5, 0.5, 0.5)$. This indicates the claim is 50% true, 50% false, and 50% indeterminate. If one interprets the numbers straightforwardly it sounds odd (summing to 1.5), but in neutrosophic terms it conveys: significant truth, significant falsehood, significant uncertainty co-exist. Such an output would signal to users that the claim is **highly controversial or ambiguous**.

The theoretical advantage here is nuance: We differentiate between *lack of evidence* indeterminacy and *conflicting evidence* indeterminacy. Both manifest as I > 0, but their combination with T and F differs. Lack of evidence: T and F are both low, I high. Conflicting evidence: T and F both high, I maybe also moderate (or can be computed from overlap as above). In the refined neutrosophic sense, these are different subcases (ignorance vs contradiction). For the scope of this paper, we treat them under the single I umbrella but conceptually note the difference.

To ensure the framework remains grounded, we align these theoretical constructs with existing fact-checking rating scales. For example, consider PolitiFact's categories [2]:

- *True* – would correspond to T ≈ 1.0, F ≈ 0, I ≈ 0.
- *Mostly True* – high T, low F, some I (perhaps minor omissions).
- *Half True* – T and F both medium, I perhaps medium (acknowledging confusion or mixed elements).
- *Mostly False* – low T, high F, some I.
- *False* – T ≈ 0, F ≈ 1.0, I ≈ 0.
- *Pants on Fire* (ridiculously false) – T=0, F=1, I=0, with maybe an extra indicator for blatant falsehood (which could be seen as an extreme F or a meta-tag).

Our neutrosophic values can emulate these with a continuous range, rather than discrete bins. This is useful because not every claim neatly fits a predefined category – sometimes it's "between Half True and Mostly True," etc. The (T,I,F) scores provide a continuum. In a practical implementation, one could map the continuum back to categories if needed for presentation (for instance, if T > 0.7 and F < 0.3, call it "Mostly True"), but the underlying availability of numeric scores means more precision and the possibility of conveying uncertainty with percentages or visuals (e.g., a tri-bar graph).

It is important to note that our framework assumes an ability to quantify support and refutation from text – an area of active research and not error-free. The methodology section will detail how we propose to calculate T, I, F in a systematic way. However, even without perfect automation, the framework could be applied by human fact-checkers as well: experts could assign subjective scores for T, I, F based on their investigation (similar to how they currently assign ratings). In that sense, this approach could unify human and machine assessments under one scheme.

To summarize the theoretical framework:

- We conceptualize fact-check verdicts as **neutrosophic triples** capturing truth, falsehood, and indeterminacy.
- These values are derived from evidence: support contributes to truth, refutation to falsehood, and lack or conflict of evidence to indeterminacy.
- The framework is flexible to allow representation of contradiction (T and F both high) and ignorance (I high with T,F low).
- It generalizes existing categorical models (binary, three-way, ratings) into a continuous tri-dimensional space.

- It is grounded in neutrosophic logic theory, ensuring that our representation has a formal semantic meaning (as per neutrosophy) and is not just an ad-hoc scoring heuristic.

Next, we detail the methodology for implementing this framework, including any formulas or algorithms for computing the neutrosophic components from data.

4. Methodology

Our methodology consists of two main parts: (1) a *literature review* that underpins the conceptual development and (2) a *computational procedure* for applying neutrosophic logic to fact-checking outcomes. The literature review process has been described in the previous section, where we synthesized findings from a range of disciplines. In this section, we focus on the computational methodology – essentially a step-by-step procedure or algorithm for deriving (T, I, F) values for a given claim using available evidence.

It is important to clarify that, at this stage, the methodology is theoretical and algorithmic; we do not yet implement it on a large scale dataset within this paper, but we outline how such an implementation *would* proceed. This approach is akin to presenting a model specification in a theoretical paper. We also provide a small hypothetical example to illustrate the calculations. Any speculative choices in the algorithm (due to lack of established solutions in literature) are explicitly noted as such.

4.1. Evidence Gathering

Input to the system: A claim C (in textual form, e.g., *"Candidate X voted for policy Y five times."*).

The first step is to gather evidence relevant to C. In an automated setting, this might involve querying search engines, databases (like Wikidata), or specialized archives (like fact-check databases or scientific literature if the claim is scientific). This step leverages information retrieval techniques. For our purposes, we assume this yields a set $E = e_1, e_2, \dots, e_n$ of evidence items. Each item e_i could be represented by a short text snippet or a document plus metadata (source name, date, etc.).

Methodological note: We are not proposing a novel IR technique here; any standard evidence retrieval approach from fact-checking research can be used. One could use the FEVER baseline retrieval or something like BM25 ranking of documents from Wikipedia. The goal is to get a reasonably comprehensive set of evidence bearing on the claim.

4.2. Evidence Classification (Stance Detection)

Each evidence item e_i is then classified with respect to the claim C into one of three categories:

- **Support:** e_i provides information that tends to verify or support the claim.
- **Refute:** e_i provides information that contradicts or undermines the claim.
- **Neutral/Irrelevant:** e_i does not directly address the claim, or the information is ambiguous/mixed.

This can be done via a stance detection model. For example, one could use a textual entailment model to check if e_i entails $C(support)$, contradicts $C(refute)$, or neither $(neutral)$ [26]. In the FEVER approach, evidence sentences carry labels "supports," "refutes," or "not enough info." Similarly, we might use a fine-tuned transformer (like a RoBERTa model trained on FEVER or related datasets) to assign these labels to each retrieved item.

Along with a categorical label, we might get a confidence score from the model. For instance, e_1 might be classified as *Support* with 90% confidence, e_2 as *Refute* with 75% confidence, etc. These confidences can serve as weights.

4.3. Aggregating Support and Refutation

We then aggregate the results of evidence classification into numeric scores $S, R, and\ U$ as defined earlier. Here is one possible formulaic approach (we label it **Method A: weighted aggregation**):

Let

$$\mathrm{s_i} = \mathrm{P(support|e_i)},$$
$$\mathrm{r_i} = \mathrm{P(refute|e_i)},$$
$$\mathrm{n_i} = \mathrm{P(neutral|e_i)}$$

as the confidence output of the stance model for each category (with $s_i + r_i + n_i = 1$). If the model is purely categorical with no probabilities, we can set $s_i = 1$ for support label and 0 otherwise, etc., but using probabilities helps smooth things.

Then define:

$$S = \frac{1}{N}\sum_{i=1}^{N} s_i,$$

$$R = \frac{1}{N}\sum_{i=1}^{N} r_i,$$

$$U = \frac{1}{N}\sum_{i=1}^{N} n_i,$$

where N is the number of evidence items considered (or potentially the number of high-quality evidence items if we impose a cutoff).

By construction, $S + R + U = 1$ because each e_i contributes one to one of the categories in a fractional way. U here represents the proportion of evidence that ended up neutral or was not found to be useful. It thus conflates two things: irrelevance and lack of evidence. One can refine this by filtering truly irrelevant results out of E in the retrieval step (only keep evidence that is potentially useful). For our purposes, U can be interpreted as "*not supporting or refuting*," which covers both evidence that is inconclusive and any gap where evidence might be missing.

Example (Hypothetical): Claim: "*The city of X spent $5 million on Y project last year.*" Evidence retrieved:

- e_1: City budget report showing $5 million allocated to Y – (Support with high confidence, say $s_1 = 0.9, r_1 = 0.1$ after entailment because perhaps it doesn't explicitly say "spent").
- e_2 : News article quoting an official denying the spending – (Refute with moderate confidence, $s_2 = 0.2, r_2 = 0.7, n_2 = 0.1$).
- e_3: A fact-check from a year ago saying "unclear how much was spent" – (Neutral or mixed, perhaps $s_3 = 0.4, r_3 = 0.4, n_3 = 0.2$ because it presents both sides).
- e_4: Unrelated article about Y project outcomes – (Neutral/irrelevant, $s_4 = 0.0, r_4 = 0.0, n_4 = 1.0$). Assume these 4 are all our evidence ($N = 4$). Then summing: $S = (0.9 + 0.2 + 0.4 + 0.0)/4 = 0.375$, $R = (0.1 + 0.7 + 0.4 + 0.0)/4 = 0.3$, $U = (0.0 + 0.1 + 0.2 + 1.0)/4 = 0.325$. We get $(T = 0.375, F = 0.3, I = 0.325)$. This would indicate the claim leans somewhat true (there is evidence for it, and somewhat less against it, but a large portion is indeterminate). A human might call that "Partly true but uncertain," which aligns.

4.4. Incorporating Conflict (Optional Extension)

If we want to allow $T + F > 1$ in cases of strong conflict (neutrosophic overset), we can introduce a conflict factor C. One simple way: define $C = \max(0, S + R - 1)$. This C will be zero unless $S + R$ exceeds 1 (meaning both support and refute are substantial). If $S + R > 1$, then we have double-counted evidence in a sense. For example, if $S = 0.8$ and $R = 0.5$, then $C = 0.3$. We can then redistribute that conflict into I`. Specifically, we could set:

- $T' = min(S, 1)$ (cap T at 1, since beyond 1 is not interpretable as probability – alternatively keep raw S, but usually S is ≤1 anyway by design).
- $F' = min(R, 1)$.
- $I' = U + C$.

So in the example of $S = 0.8, R = 0.5$: originally $U = (1 - S - R)$ if normalized = not valid here since $S + R > 1$; using conflict, say originally we had $U = 0$ in that scenario (maybe no neutrals, just conflicting sources). Then $C = 0.3$, we set $I' = 0.3$. Now $T' = 0.8, F' = 0.5, I' = 0.3$. This sums to 1.6, reflecting an overspecification due to conflict, which in neutrosophic logic is acceptable [9].

This method of capturing conflict is **speculative** – it is one way to encode overlapping truth and falsehood evidence, inspired by neutrosophic "*overlap*" allowance. Alternatively, one could keep normalization but simply note high values of both S and R as indicating conflict. For simplicity, the main methodology we present will stick to normalized S, R, U for computing (T, I, F),and we will treat conflict analysis as a qualitative add-on in discussion.

4.5. Outputting Neutrosophic Scores

After computing S, R, U (and optionally C), we produce the neutrosophic truth-value for claim C:

$$\text{Score}(C) = (T, I, F),$$

with:

- T = S (support degree),
- $F = R$ (refute degree),
- $I = U$ (uncertainty degree),
- (If conflict considered, $I = U + C$).

This score can then be mapped or interpreted in various ways:

- As a triple of numbers (e.g., in a data output or an API).
- As a visualization (like a triangular radar chart or three thermometers).
- As a textual summary: for example, use thresholds to say "mostly true" if T is dominant but I not too high, or "*uncertain*" if I is high, etc., possibly accompanied by an explanation like "*Evidence is mixed: approximately X% supports and Y% refutes the claim.*"

The methodology up to this point is the *computational core*. We will now address some methodological considerations and validity:

Data Requirements: This approach assumes a pool of evidence. If none is found ($N = 0$), we default to $(T = 0, F = 0, I = 1)$ – completely indeterminate. This is sensible: the claim is unverified. In practice, $N = 0$ would mean our search missed or there truly is no info. Perhaps a threshold of minimal evidence should be defined where any extremely low N yields a default of high I.

Calibration: The interpretation of the scores depends on how well the stance detection works. If the stance model is noisy, it could mis-classify neutral evidence as support or refute, which would skew T or F. In a rigorous implementation, one would calibrate the model or manually vet the evidence classification for important claims. Our framework could also be used in a hybrid human-AI fashion: the AI suggests a (T, I, F) but a human fact-checker can adjust the values if needed based on their expert reading of evidence.

Transparency: One nice aspect is that the final (T, I, F) can be traced back to evidence contributions. This is important for an explainable system. We imagine an interface where a user sees "Claim X: 0.7 true, 0.2 indeterminate, 0.1 false" and can click to see the evidence that supported or refuted it. This aligns with calls for explainable fact-checking [27].

Systematic Literature Integration: The above methodology was developed after analyzing how humans do fact-checking (collect evidence, weigh pro and con) and how uncertainty methods (fuzzy, D-S, etc.) combine evidence. It is a novel synthesis; we did not find prior algorithms that *directly* produce a neutrosophic triple for claim verification, which is why parts of this method are necessarily *proposed* rather than validated. We mark this as an area where future evaluation is needed – for

example, testing this on a dataset like AVeriTeC [7] to see if the neutrosophic scores correlate well with the ground truth labels (Supported, Refuted, NEI, Conflicting).

In summary, the methodology provides a blueprint for computing neutrosophic truth values from multiple evidence inputs. It leverages existing NLP tasks (stance detection) and adds a novel aggregation logic to produce the triple score. Next, we present the proposed framework in a more holistic way, situating this computation in the context of a fact-checking system and discussing how it addresses the goals of quantifying indeterminacy.

5. Proposed Framework

Building on the theoretical and methodological foundations, we now outline the proposed **Neutrosophic Fact-Checking Framework**. The framework can be thought of as an enhancement layer for automated fact-checking systems, enabling them to output neutrosophic truth-value assessments. It consists of the following components or stages:

1. **Claim Processing and Understanding:** The system first parses the claim to understand its meaning and scope. This may involve entity recognition (who or what is being discussed), temporal cues (does it refer to "*last year*"?), and so on. This stage ensures that evidence retrieval can be focused correctly. (For instance, identifying that the claim involves "city of X" and "project Y" and a money amount will guide the retrieval to city budget documents or news about project Y.)
2. **Evidence Retrieval:** Using the processed claim, the system performs queries against various information sources: news articles, fact-check databases, official records, etc. It may use multiple search strategies (keyword search, leveraging knowledge graph relations, using claim embeddings to find similar cases). The output is a set of relevant documents or snippets.
3. **Evidence Evaluation (Stance Detection):** Each retrieved item is evaluated for its stance relative to the claim (supporting, refuting, neutral). This uses NLP models as described in the Methodology. Importantly, the system might also assess source reliability here; evidence from a highly reputable source might be weighted more. (This could be implemented by multiplying s_i and r_i by a source reliability factor between 0 and 1, for example.)
4. **Aggregation into Neutrosophic Values:** The core of the framework takes the stance-labeled evidence set and computes aggregate T, I, F values. This is done according to the formulas and procedures outlined earlier (e.g., averaging support and refute scores, computing the indeterminacy as the remainder or including conflict adjustments). The result is a triple (T, I, F) for the claim.
5. **Decision/Action Module:** Depending on the application, the neutrosophic assessment can be used in different ways:
    - If it's a public fact-checking tool, it might directly display the (T,I,F) to users with an explanation.
    - If it's for aiding human fact-checkers, it might trigger certain actions: e.g., if I is high because of lack of information, it could suggest areas for further research; if I is high due to conflict, it could flag the claim as requiring expert intervention to resolve.
    - If integrated into content moderation (e.g., social media platforms), thresholds could be set: a claim with very high F might be flagged or labeled false, one with high I might be labeled "unverified" rather than taking it down, reflecting uncertainty rather than confirmed falsehood.
6. **Explanation Interface:** Because transparency is vital, the framework includes the ability to explain how it arrived at (T, I, F). This means linking back each component to evidence:
    - For Truth component T: show the top supporting evidence sources.
    - For Falsehood component F: show the top refuting evidence.
    - For Indeterminacy I: explain why uncertainty remains – e.g., "No sources with information on XYZ were found" (ignorance) or "Sources disagree on this fact" (conflict) or "The claim is too vague to verify precisely" (ambiguity).

This explanatory step is crucial to maintain trust in the system's outputs and allows users (or fact-checkers) to validate or override the conclusions.

The framework can be visualized as a flowchart: Claim -> [Retrieve] -> Evidence set -> [Classify stances] -> Labeled evidence -> [Aggregate] -> (T,I,F) -> [Output & Explain].

To illustrate how the framework works in practice, consider an example scenario (using a hypothetical claim that resembles real-world cases):

Claim: "*COVID-19 vaccines cause infertility in women*." (This is a misinformation claim that circulated.)

- *Claim Processing:* Identify key elements: "COVID-19 vaccines" as subject, "cause infertility in women" as the assertion to check. The claim is causal and medical.
- *Evidence Retrieval:* The system queries medical studies, public health statements, fact-check articles. It finds:
    - A CDC report saying there's no evidence of infertility caused by vaccines.
    - A viral Facebook post (unreliable source) where someone claims she became infertile after a vaccine.
    - A fact-check from a reputable site debunking this infertility claim.
    - Perhaps no scientific study supporting the claim (absence of evidence in journals).
- *Evidence Evaluation:*
    - The CDC report: stance = Refutes (with high confidence) – it explicitly says "no evidence of infertility".
    - The viral post: stance = Supports (low confidence, since it's anecdotal, but it is a piece of evidence used by claim proponents).
    - The fact-check article: stance = Refutes (high confidence, presumably).
    - The lack of any scientific study *could* be considered evidence of absence; more explicitly, maybe an expert statement: "Experts say infertility is not a known side effect" – also Refute. So we might end up with many refuting evidence pieces, maybe one or two supporting (mostly from dubious sources), and a lot of neutral scientific literature that just doesn't mention the issue (which counts toward U).
- *Aggregation:* Suppose after weighting, we get roughly $S = 0.1$ (only fringe support), $R = 0.8$ (strong refutation from credible sources), $U = 0.1$ (some uncertainty or simply the remainder). The conflict measure might be low because support is very low relative to refute. So $(T = 0.1, I = 0.1, F = 0.8)$.
- *Decision:* The system would label this claim as false (since F is dominant), but also note a small I (perhaps acknowledging that ongoing studies are still collecting long-term data, hence not absolutely zero uncertainty – we leave a bit of I).
- *Explanation:* It would present: "False (80% false, 10% true, 10% uncertain)". Then list evidence: 5 studies/health org statements that found no link (for falsehood evidence), perhaps one anecdotal claim that alleged infertility (for truth evidence, though from a low-quality source), and mention "scientific consensus indicates no causal link, while no concrete evidence supports the claim". The 10% uncertainty could be explained as "No long-term study can *completely* rule out all effects, but currently no indication of this effect."

This example shows the framework providing a quantified and explainable outcome, aligning with how a fact-checker might communicate: "*There's **no proof** and strong evidence against this claim, so it's rated false*." The neutrosophic twist is that we can explicitly attach numbers to that statement.

One might ask: why not just output a single probability like "there's an 80% chance this claim is false"? The answer is subtle but important. A single probability (say $P(True) = 0.1$ in the above) doesn't distinguish between lack of evidence and presence of counter-evidence. Our triple does – we see both F high and I low, meaning it's not just unknown, it's actively disproved. If instead the claim was "New vaccine X cures cancer" (totally unsubstantiated, little data yet), we might get $T \sim 0$ (no evidence of truth), $F \sim 0$ (no evidence against, because it's just unknown), $I \sim 1$ (no one really knows)

– which is a very different scenario from the infertility case. A single probability might also be near 0 for "cures cancer", but the *meaning* is different. Neutrosophic output gives that nuance.

Framework Validation: The proposed framework is admittedly complex, and validating it would require careful experimentation:

- We could test it on existing fact-check datasets. For each claim, compute (T, I, F) from the evidence and compare with the ground-truth rating. Do claims labeled "True" get T high, "Half True" get moderate T and F, etc.? This would help fine-tune the interpretation.
- Conduct user studies: Do journalists or end-users find the neutrosophic outputs helpful or clearer than a simple true/false? This is more of a communication question but central to the framework's utility. Prior research suggests people have trouble with probabilities, but presenting three numbers might be even more challenging unless done carefully. This is discussed further in the next section.
- Edge cases: The framework should be robust to claims that are opinion-based or value judgments (which might yield all neutral evidence and thus come out (0,1,0) – effectively "unverifiable"). It should be able to identify those as not factual claims at all, possibly. Integrating a claim type classifier could be an extension (to first decide "Is this claim fact-checkable?").

Finally, while the focus is on machine fact-checking, the framework could aid human fact-checkers as a structured way to think about their conclusion. In fact, a fact-checker could manually assign T,I,F after reviewing evidence, to avoid strict category boundaries. This might even improve inter-rater consistency if guidelines are given (for example, PolitiFact could define that "Half True" corresponds roughly to T and F both in 0.3–0.7 range, etc.). However, these suggestions are **speculative** – they would require testing in practice with fact-checking teams.

6. Discussion

The introduction of a neutrosophic approach to fact-checking brings several potential advantages, as well as challenges and open questions. In this section, we discuss the implications of the proposed framework, examine how it compares to or complements existing methods, highlight limitations (including speculative aspects that need further research), and outline future directions.

Beyond single claim veracity, this framework could be extended to analyze broader concepts like media bias or journalistic neutrality. A high 'I' score, for instance, might represent balanced reporting in some contexts, but careful operationalization, potentially using refined indeterminacy, would be needed to distinguish constructive neutrality from vagueness or false balance.

Another promising direction is exploring dynamic neutrosophic models. Since the evidence regarding a claim can change over time, representing the truth value as functions T(t), I(t), F(t) could allow tracking the evolution of certainty and indeterminacy as a story develops or new information emerges.

6.1.Contributions and Advantages

1. Enhanced Uncertainty Quantification: The primary benefit of our framework is the explicit quantification of uncertainty (indeterminacy). Rather than a black-box confidence score or a vague "unproven" label, we assign a value to *how indeterminate* a claim's truth status is. This directly addresses the problem identified in journalism research that many claims cannot be fully verified or falsified [1]. By quantifying indeterminacy, we make the system's humility transparent – it can say, in effect, "this portion of the truth value is undetermined." This could improve user trust in automated fact-checkers: users may be more accepting of an AI's conclusion if it also honestly conveys uncertainty, similar to how weather forecasts include a probability of rain rather than a yes/no prediction.

2. Nuanced Representation of Partial Truth: Our neutrosophic method can naturally represent partial truths, which are common in news [3]. For example, a claim that is technically correct but misses context might get a moderate T, some F (for the misleading aspect), and moderate I (for the

unresolved context). This is more informative than just labeling it "Half True" without detail. It aligns with the way fact-checkers write narrative explanations for half-true claims – now those nuances are captured in the model's output. By capturing degrees of truth and falsehood simultaneously, the system can handle "shades of gray" in a principled manner.

3. Handling Contradictory Evidence: Traditional systems struggle with contradictory evidence because they force a single verdict. In our framework, contradictory evidence simply increases both T and F, potentially also I, indicating disagreement [9]. This is a more faithful reflection of reality when sources clash. A concrete scenario: Claim "Effectiveness of Drug A is X%" – one study says 50%, another says 70%. An AI might pick one as more reliable and say "True (70%)" whereas our system could output T0.7, F0.7 (contradiction) and I a bit to account for unknown factors. This tells the user the evidence is conflicting. It mirrors how a thorough fact-check might say "experts are divided". No mainstream computational model currently expresses such divisions except by maybe issuing two separate fact-checks. Our single triple captures it succinctly.

4. Decision-making Flexibility: The neutrosophic output could feed into **decision systems** with adjustable thresholds. Platforms could decide: if F is above 0.7 and I is low, flag as false; if I is very high (say >0.8) and both T,F are low, label as "unverified claim – proceed with caution"; if T and F are both moderate and I moderate, label as "disputed/controversial". This is similar to three-way decisions but with more gradation. It allows policies that are more nuanced than either leaving content up or removing it. For instance, social media might not want to remove a claim that's not clearly false (high I), but they might want to tag it as "unverified". Our framework provides a quantifiable basis for such actions.

5. Alignment with Human Reasoning: Neutrosophic logic, as noted, is said to be *"highly integrated with how humans think, where an indefinite environment is [our] ordinary space to draw a conclusion"* [24]. In fact-checking, human experts often think in terms like: "We have strong evidence for this part, but there's still some doubt here, and there's a piece of contrary evidence there." Our model's triple reflects that style of reasoning. In essence, it formalizes an expert's mental checklist: pro evidence, con evidence, unknowns. This could potentially make it easier for experts to interact with or override the system – they could tweak T,I,F themselves if they disagree, giving them a structured way to provide feedback.

6.2. Challenges and Limitations

1. Interpretability: While the triple is rich, it might be cognitively challenging for end-users to interpret three numbers. Some users might be confused by seeing "Truth=0.4, False=0.4". Does that mean the claim is both true and false? This runs counter to classical logic and could be misinterpreted. Thus, careful user interface design is needed. We might convert the triple into a verbal statement or a visualization (like a pie chart with overlapping slices for true and false – a non-standard concept). Educating the public or journalists in understanding such output is non-trivial. This is a trade-off between nuance and simplicity. As a solution, we might keep the triple internal or for expert users, and still present simpler labels outwardly. The triple can inform a more conventional rating that is shown. This way, the heavy lifting is done neutrosophically, but communication is done in familiar terms (similar to how weather models are complex but output "20% chance of rain" which is easy to grasp).

2. Reliability of Stance Detection: The framework's accuracy is bound by the performance of sub-components like evidence retrieval and stance classification. If the system misses key evidence or misclassifies neutral text as support, the (T, I, F) will be off. For instance, propaganda websites might be over-represented in search results, providing a lot of "supporting" evidence for a false claim, which could trick the system into a high T unless source reliability is accounted for. We addressed this by suggesting weighting by source credibility, but that introduces its own challenges (identifying credibility, not reinforcing biases, etc.). This is not a problem unique to our framework, but our output might be seen as more authoritative due to numeric precision, so errors could be misleading. A wrong neutrosophic assessment might be more dangerous if users take the numbers at face value. Therefore,

ensuring high-quality evidence processing is crucial, and any deployment should have rigorous evaluation against known ground truths.

3. Lack of Ground Truth for Indeterminacy: In supervised machine learning, we can train a model to predict true/false if we have labeled data. But we rarely have data labeled with an indeterminacy degree. Fact-check verdicts like "Unproven" or "Half True" are somewhat subjective and categorical. Mapping them to an exact I value or partition of T/F is not straightforward. Our framework is currently hand-designed rather than learned; it uses an algorithmic aggregation approach. This is reasonable for a first attempt, but it may need calibration. For example, do humans agree that a given claim is $(0.4,0.4,0.2)$? We might only know humans said "Half True". There could be multiple (T,I,F) combinations that correspond to "Half True". So evaluating the correctness of our triple is tricky. It might require asking experts to provide such triples as ground truth, which is a new kind of annotation task. Until then, part of our model remains *speculative* in that we assume our method produces meaningful triples, but they aren't directly validated against human judgment except qualitatively. This is an area for future empirical study.

4. Speculative Elements: Some elements of our method, like the conflict adjustment (C) and even the linear weighting of evidence, are not drawn from prior fact-checking literature but rather inspired by neutrosophic theory and analogous methods. These are **speculative** in the sense that they haven't been proven effective in this domain yet. It's possible that a simpler approach (like just normalizing as we did initially) is sufficient and adding C complicates things without clear benefit. We flagged these as speculative in the methodology. Going forward, an iterative approach would test variations:

- Without conflict overlap (strict normalization).
- With conflict considered.
- Perhaps with other refinements (like splitting I into I1 for ignorance vs I2 for contradiction as Patrascu (2016) suggests [25]). We would need to see which correlates best with expert evaluations or which is easiest to interpret.

5. Communicating Indeterminacy in Social Context: In the misinformation ecosystem, labeling something uncertain or indeterminate can be double-edged. On one hand, it's truthful to say "we aren't sure". On the other hand, purveyors of false claims might exploit any uncertainty label to say "See, they can't prove me wrong!" For example, early in a pandemic, many claims might be unproven simply because science hasn't caught up; labeling them indeterminate (while accurate) might lead some audiences to give them more credence than deserved (thinking "it could be true"). Fact-checkers often have to tread carefully here – sometimes they'll phrase as "No evidence for X" rather than "We don't know", to avoid implying a 50-50 chance when it's a baseless claim. Our framework would label a baseless claim as (0, ~1, 0) which literally says "no evidence either way". This could be misread as "*maybe true, maybe not*". So there is a communications challenge: how to convey that something with I=1 and T=0 is more of a "lack of any supporting evidence" (implying likely false until evidence appears) vs something with I=1 and moderate T and F is "scientifically unresolved debate". Thus the context and phrasing around indeterminacy will matter. Perhaps we need sub-labels even in presentation: "unsubstantiated" vs "unclear" vs "conflicting evidence", etc., which correspond to different flavors of indeterminacy. Indeed, implementing refined neutrosophic categories might help here (distinguish ignorance from conflict). This is a subtlety to explore in future refinements.

6.3. Comparison with Related Approaches

It's useful to situate our proposal relative to other frameworks:

- **Dempster-Shafer in Fact-Checking:** To our knowledge, few have applied D-S theory to fake news detection, though it could be. D-S would allow an *Uncertainty mass*. Our approach is somewhat analogous but arguably more direct for this application, because we explicitly identify evidence that supports or refutes and allocate "unknown" to what's not covered. D-S combination rules could be applied if we had independent evidence sources; that might be a variant of our method (we did a simple average, D-S would do something else). Possibly a fruitful comparison in future work: implement a D-S fusion of evidence and compare its

output (belief intervals) to our neutrosophic output. We suspect they would correlate strongly, as both handle unknowns explicitly.

- **Three-Way Decisions:** Our model can be seen as a soft version of three-way decisions [18]. Instead of a hard accept/reject/defer, we give degrees to each. It's like fuzzy three-way decisions. Indeed, one could derive a three-way decision by looking at which of T, F, or I is largest: if T > F and T > I, then accept (claim likely true); if F is greatest, reject (claim likely false); if I is greatest, defer (cannot decide). That would yield a decision identical to FEVER's categories possibly. But the values provide more information than just the argmax. Thus, our approach generalizes three-way decision-making. Conversely, three-way decision theory could provide a theoretical validation for having that third category – something we've assumed as useful. There is a body of work on optimization of when to defer a decision that could be applied to tune our I threshold.
- **Quantitative Journalism Research:** There have been attempts to quantify media credibility or bias on numeric scales. For example, some work measures *credibility score* of news articles via supervised models. However, quantifying the truth content with a triple is novel. One related concept is the idea of a "truth meter" or "truth score". For instance, computation journalism studies by Wang (2017) who built the LIAR dataset attempted to automatically classify statements into the 6 PolitiFact categories [28]. They essentially train a classifier to output those labels. Our approach is different in that we propose constructing the value via evidence rather than classification, which arguably makes it more explainable and adaptable to new claims without similar training data. Another relevant concept is **stance detection + verdict** that is used in FEVER-like systems: they output "Supported"/"Refuted" along with evidence. Our framework extends that by adding a degree and the middle state.
- **Neutrosophic Applications:** Within neutrosophic literature, others have used it for things like sentiment (as cited) or multi-criteria decisions. Those often involve designing membership functions for T, I, F for a particular problem context [24]. In our case, we effectively designed membership functions based on evidence counts/weights. We also considered something analogous to *single-valued neutrosophic sets* (we kept values in [0,1]). Another variant is *interval neutrosophic* where instead of a number you give an interval for each of T,I,F to express uncertainty about the values themselves [19]. We did not go that far; it would be like saying "T is between 0.3 and 0.5" to also account for model uncertainty. That would be even more complicated to present. For now, point estimates seem sufficient.

6.4. Future Directions

This research opens up several avenues for follow-up:

- **Empirical Evaluation:** The most pressing next step is to implement the framework on a representative set of claims (perhaps from a fact-checking dataset) and evaluate its outputs. How often do humans agree with the (T,I,F) assessment? We could enlist fact-checkers to rate claims on a 0-1 scale for truth and falsehood, or to allocate 100 points among True/False/Unknown for a claim, then compare with our system. This would provide validation and also highlight systematic biases (e.g., maybe our system tends to overestimate I for lack of evidence, whereas experts might be comfortable calling something false instead of unknown in some contexts).
- **Refinement of Indeterminacy Types:** As noted, integrating the idea of refined neutrosophic components[25] could improve how we communicate uncertainty. We could attempt to automatically distinguish *why* I is high. For instance, if a claim is very specific and factual but we just didn't find evidence, that's ignorance-type. If a claim is vague or value-laden, that's a kind of indeterminacy due to *unclear claim definition*. If claim has supporting and opposing evidence, that's contradiction-type. The framework could output not just I value, but I_type. This is speculative, but potentially useful. It might require some rule-based inference or additional NLP (like detecting vagueness in the claim, or checking if both T and F are above some threshold indicates contradiction).

- **User Interface and Communication Research:** Collaborations with communication scientists and designers could figure out the best way to present neutrosophic fact-check results to the public. This could involve user testing of different representations (numbers vs verbal vs charts). It might turn out that a simplified message is needed, in which case we ensure the framework's richness is utilized under the hood for system decision-making, even if not all of it is exposed to the user.
- **Integration with AI Explanation Systems:** There's increasing demand for explainable AI, especially as large language models (LLMs) are being considered for fact-checking aids [29,30]. Our approach could be complementary: an LLM could be prompted to find evidence and our system could formalize the scoring. Or vice versa, our system produces (T,I,F) and an LLM is prompted to produce a concise narrative explanation given those scores and evidence (ensuring consistency). There is fertile ground for combining symbolic/neutrosophic reasoning with the flexible language abilities of LLMs.
- **Cross-Domain Generality:** While we focused on news, the approach could apply to scientific claims, rumors, even legal evidence assessment. Essentially any claim where evidence can be compiled could be evaluated in this triadic way. It would be interesting to apply it to a dataset like climate change claims, health myths, etc., and see if it robustly handles different domains. The weights might need adjusting if evidence quality varies by domain.
- **Dynamic Updates:** Facts change over time (especially in science or developing news stories). A claim's truth-value might shift from indeterminate to known, or even flip from false to true if new evidence emerges (though rare, it happens). A neutrosophic system could be updated dynamically with new evidence, and its output would change in a measurable way. Tracking the trajectory of (T,I,F) over time could provide a novel way to visualize how a story evolves (e.g., early on I is high, later T grows as evidence accumulates). This temporal aspect would be an intriguing subject for further study, bridging fact-checking with epistemic monitoring.

6.5. Ethical and Social Considerations

It's worth noting some ethical considerations briefly: Providing quantified truth assessments has a risk of conveying unwarranted certainty or authority. Our framework must be used responsibly. Fact-checking itself can become politicized; a tool that labels something with numbers might be attacked as biased or "who gave you the authority to assign these numbers?". Transparency about methodology (which we advocate by explaining evidence contributions) is key to mitigating that. Additionally, one must consider the potential misuse: could someone input false evidence to skew the output? In an automated system, adversarial actors might game the evidence retrieval (e.g., SEO tactics to flood search results with supporting junk for a false claim, making the AI think T is higher). This again circles back to source credibility assessment and possibly manual curation for important topics. These issues are not unique to neutrosophic fact-checking but are heightened by the quantitative nature of the output.

In conclusion, the neutrosophic approach to quantifying indeterminacy in news fact-checking is a promising theoretical development that addresses a clear gap: the need to represent uncertainty in automated claim verification. Our framework offers a novel synthesis of ideas from journalism, computational linguistics, and neutrosophic logic. It contributes a structured method to capture *"truthiness"* in a multi-dimensional way, pushing beyond the conventional binary or tiered verdicts. While much work remains to implement and validate this approach, it lays conceptual groundwork for more transparent and nuanced AI fact-checkers – tools that not only answer whether a claim is true or false, but also convey *how well* it is supported and *how much* we do not know.

7. Conclusions

We presented a comprehensive exploration of applying **neutrosophic logic** to the domain of automated news fact-checking. Through an interdisciplinary literature review, we established that traditional fact-checking outcomes often reside in a gray area of partial truths and uncertainties [1,3], and current automated approaches lack a robust mechanism to quantify this indeterminacy. To

address this gap, we developed a theoretical framework that represents fact-check verdicts as neutrosophic triples (T, I, F), denoting degrees of truth, indeterminacy, and falsehood. We proposed a methodology for deriving these values from evidence, using natural language processing techniques for stance detection and drawing inspiration from neutrosophic theory to handle contradictory and incomplete information.

Our framework is a novel contribution that bridges the gap between the binary logic of computing and the nuanced reasoning of journalism. It allows an automated system to output statements like, "*Claim X is 70% true, 0% false, and 30% indeterminate*," in contrast to a simplistic true/false label. This granular outcome aligns with the multi-valued ratings used by fact-checkers (True, Mostly True, Half True, etc.) but provides even more flexibility by quantifying the extent of truth and uncertainty [2]. We showed through examples and discussion that such outputs can capture the state of evidence: whether a claim is well-supported, hotly disputed, or simply lacking research.

The potential benefits of this approach include improved transparency of AI fact-checking systems, better communication of uncertainty (which is ethically important to avoid overstating what is known), and enhanced decision-making capabilities for platforms dealing with misinformation. By explicitly modeling indeterminacy, our method can identify claims that merit further investigation or cautious treatment, rather than forcing a premature verdict. Furthermore, it creates opportunities for hybrid human-machine workflows, where the machine provides a structured assessment and human experts can refine or explain it.

However, we also acknowledge several challenges. The concept of a claim being simultaneously partly true and partly false is unconventional and may face acceptance hurdles without careful user experience design. The accuracy of the neutrosophic scores is contingent on the quality of evidence analysis; thus, advances in information retrieval and stance detection are complementary to the success of our framework. We also flagged that parts of our approach, particularly the exact numeric mappings and the treatment of conflicting evidence, are **speculative and would benefit from empirical calibration and evaluation:** implementing the framework and testing its performance on diverse, real-world fact-checking datasets (e.g., FEVER, AVeriTeC). This involves comparing the generated (T, I, F) scores against existing ground-truth labels and, ideally, against human expert judgments specifically elicited using a similar triadic scale. Such evaluation will be crucial for refining the aggregation algorithms and calibrating the model.

In terms of limitations, this work has been theoretical. Further development should focus on refining the representation of indeterminacy, potentially incorporating distinctions between ignorance (lack of evidence), contradiction (conflicting evidence), and ambiguity (vague claims), possibly drawing on refined neutrosophic set concepts. Significant effort must also be dedicated to user interface and communication research to determine the most effective ways to present these nuanced outputs to journalists, fact-checkers, and the general public, ensuring clarity over confusion. Exploring the integration with large language models (LLMs) offers another promising avenue, potentially leveraging LLMs for improved evidence gathering or for generating natural language explanations based on the calculated neutrosophic scores. Investigating the framework's cross-domain applicability and developing capabilities for dynamic updating of scores as evidence evolves over time represent further valuable extensions

We lay the groundwork for a new approach to automated fact-checking, one that *quantifies uncertainty* rather than eliding it. We have shown through reasoned argument and reference to existing research that neutrosophic logic provides a philosophically and practically sound way to represent the indeterminate nature of many news claims [9,19]. Our proposed framework is an initial step toward implementing this theory in computational systems. By marrying the precision of computation with the nuance of neutrosophy, we aim to contribute to the development of fact-checking tools that are not only accurate but also honest about the limits of what can be determined. Such tools will be increasingly valuable in a world awash with information and uncertainty.

Quantifying indeterminacy in news through a neutrosophic method offers a pathway to more resilient and trustworthy fact-checking processes – ones that can withstand the complexities of real-world information and help both AI and humans navigate the murky waters between true and false.

While substantial work remains in implementation, validation, and refinement, this research lays the conceptual groundwork for a new generation of automated fact-checking systems. By embracing and quantifying indeterminacy through the lens of neutrosophic logic, we move towards tools that are not only computationally powerful but also more epistemically honest about the inherent complexities and uncertainties in assessing real-world information. The ultimate goal is to contribute to more resilient, trustworthy, and nuanced fact-checking processes capable of navigating the challenging information landscape.

Funding: This research received no external funding

Conflicts of Interest: The authors declare no conflict of interest

References

1. Graves, L. Understanding the Promise and Limits of Automated Fact-Checking 2018.
2. Li, J.; Chang, X. Combating Misinformation by Sharing the Truth: A Study on the Spread of Fact-Checks on Social Media. *Inf Syst Front* **2022**, 1–15, doi:10.1007/s10796-022-10296-z.
3. Sandeep, S.; Bhattacharyya, P. Detecting and Debunking Fake News and Half-Truth: A Survey. **2023**.
4. Vosoughi, S.; Roy, D.; Aral, S. The Spread of True and False News Online. *Science* **2018**, *359*, 1146–1151, doi:10.1126/science.aap9559.
5. Guo, Z.; Schlichtkrull, M.; Vlachos, A. A Survey on Automated Fact-Checking 2022.
6. Thorne, J.; Vlachos, A.; Christodoulopoulos, C.; Mittal, A. FEVER: A Large-Scale Dataset for Fact Extraction and VERification 2018.
7. Schlichtkrull, M.; Guo, Z.; Vlachos, A. AVeriTeC: A Dataset for Real-World Claim Verification with Evidence from the Web. In Proceedings of the Advances in Neural Information Processing Systems; Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S., Eds.; Curran Associates, Inc., 2023; Vol. 36, pp. 65128–65167.
8. Paraskevas, A.; Smarandache, F. Modeling Social Evolution, Involution, and Indeterminacy: A Neutrosophic-Cultural Algorithm Approach. *Neutrosophic Sets and Systems* **2025**, *77*.
9. Mathematics, Physics and Natural Sciences Division,The University of New Mexico, 705 Gurley Ave., Gallup, NM 87301, USA; Smarandache, F. Indeterminacy in Neutrosophic Theories and Their Applications. *IJNS* **2021**, doi:10.54216/IJNS.150203.
10. Smarandache, F. The Rank Order of the Neutrosophic Triplets (T, I, F) May Change after Normalization. *International Journal of Research in Industrial Engineering* **2024**, *13*, 427–435, doi:10.22105/riej.2024.487239.1489.
11. Jr, E.C.T.; Lim, Z.W.; Ling, R. Defining "Fake News." *Digital Journalism* **2018**.
12. Allcott, H.; Gentzkow, M. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* **2017**, *31*, 211–236, doi:10.1257/jep.31.2.211.
13. Hassan, N.; Arslan, F.; Li, C.; Tremayne, M. Toward Automated Fact-Checking: Detecting Check-Worthy Factual Claims by ClaimBuster. In Proceedings of the Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Association for Computing Machinery: New York, NY, USA, August 13 2017; pp. 1803–1812.
14. Zadeh, L.A. Fuzzy Sets. *Information and Control* **1965**, *8*, 338–353, doi:10.1016/S0019-9958(65)90241-X.
15. Atanassov, K.T. Intuitionistic Fuzzy Sets. *Fuzzy Sets and Systems* **1986**, *20*, 87–96, doi:10.1016/S0165-0114(86)80034-3.

16. Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press, 1976; ISBN 978-0-691-10042-5.
17. Yao, Y. Three-Way Decisions with Probabilistic Rough Sets. *Information Sciences* **2010**, *180*, 341–353.
18. Zhang, X.; Ouyang, T. Granular Description of Uncertain Data for Classification Rules in Three-Way Decision. *Applied Sciences* **2022**, *12*, 11381, doi:10.3390/app122211381.
19. Zhang, H.; Wang, J.; Chen, X. Interval Neutrosophic Sets and Their Application in Multicriteria Decision Making Problems. *ScientificWorldJournal* **2014**, *2014*, 645953, doi:10.1155/2014/645953.
20. El-Hefenawy, N.; Metwally, M.A.; Ahmed, Z.M.; El-Henawy, I.M. A Review on the Applications of Neutrosophic Sets. *Journal of Computational and Theoretical Nanoscience* **2016**, *13*, 936–944, doi:10.1166/jctn.2016.4896.
21. Smarandache, F. *A UNIFYING FIELD IN LOGICS: NEUTROSOPHIC LOGIC. NEUTROSOPHY, NEUTROSOPHIC SET, NEUTROSOPHIC PROBABILITY AND STATISTICS*; 6th ed.; InfoLearnQuest: Michigan, USA, 2007; ISBN 978-1-59973-892-5.
22. Smarandache, F. Introduction to Neutrosophic Statistics 2014.
23. Smarandache, F. Introduction to Neutrosophic Measure, Neutrosophic Integral, and Neutrosophic Probability. *Branch Mathematics and Statistics Faculty and Staff Publications* **2013**.
24. Essameldin, R.; Ismail, A.A.; Darwish, S.M. Quantifying Opinion Strength: A Neutrosophic Inference System for Smart Sentiment Analysis of Social Media Network. *Applied Sciences* **2022**, *12*, 7697, doi:10.3390/app12157697.
25. Patrascu, V. Refined Neutrosophic Information Based on Truth, Falsity, Ignorance, Contradiction and Hesitation. *Neutrosophic Sets and Systems* **2016**, *11*.
26. Liu, Y.; Zhu, C.; Zeng, M. Modeling Entity Knowledge for Fact Verification. In Proceedings of the Proceedings of the Fourth Workshop on Fact Extraction and VERification (FEVER); Aly, R., Christodoulopoulos, C., Cocarascu, O., Guo, Z., Mittal, A., Schlichtkrull, M., Thorne, J., Vlachos, A., Eds.; Association for Computational Linguistics: Dominican Republic, November 2021; pp. 50–59.
27. Warren, G.; Shklovski, I.; Augenstein, I. Show Me the Work: Fact-Checkers' Requirements for Explainable Automated Fact-Checking 2025.
28. Wang, W.Y. "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. In Proceedings of the Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers); Barzilay, R., Kan, M.-Y., Eds.; Association for Computational Linguistics: Vancouver, Canada, July 2017; pp. 422–426.
29. Fu, Y.; Guo, S.; Hoffswell, J.; Bursztyn, V.S.; Rossi, R.; Stasko, J. "The Data Says Otherwise"-Towards Automated Fact-Checking and Communication of Data Claims. In Proceedings of the Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology; October 13 2024; pp. 1–20.
30. Quelle, D.; Bovet, A. The Perils & Promises of Fact-Checking with Large Language Models. *Front. Artif. Intell.* **2024**, *7*, 1341697, doi:10.3389/frai.2024.1341697.

Received: Jan 9, 2024. Accepted: July 5, 2025