E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

***IANNA JOURNAL OF INTERDISCIPLINARY STUDIES*, Vol. 8, No. 2 (June 2026).**
**Journal's website: https://iannajournalofinterdisciplinarystudies.com**
**Article History: Received 20 January 2026; Revised, 20 February 2026; Accepted, 22 March 2026**

# The Paradox of Fluency: A Comparative Analysis of Traditional and Neural Machine Translation Systems through an Ecological Lens

**Oleksandra Borysenko**[1*]
https://orcid.org/0000-0001-9138-6612
**Oksana Dubrova**[2]
https://orcid.org/0000-0002-8150-1215
**Artur Gudmanian**[3]
https://orcid.org/0000-0002-4196-2279
**Svitlana Sokolovska**[4]
https://orcid.org/0000-0002-2335-1765
**Oksana Kotenko**[5]
https://orcid.org/0009-0000-3350-350X

[1]Foreign Languages Department, Faculty of Economics, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine.
[2]Foreign Languages Department, Faculty of Management, Logistics and Tourism, National Transport University, Kyiv, Ukraine.
[3]Department of English, State University of Information and Communication Technologies, Kyiv, Ukraine.
[4]Department of German Philology and Foreign Literature, Educational and Scientific Institute of Foreign Philology, Zhytomyr Ivan Franko State University, Zhytomyr, Ukraine.
[5]Department of Ukrainian Philology and Foreign Literature, Faculty of Social Sciences and Humanities and Law, Bogdan Khmelnitsky Melitopol State Pedagogical University, Zaporizhzhia, Ukraine

*Correspondence author email: academia_ua@ukr.net

**Abstract**

**Background:** The rapid evolution of Neural Machine Translation (NMT) has produced unprecedented linguistic fluency; however, this progress has intensified ethical and ecological dilemmas regarding model interpretability, sustainability, and cross-domain stability. As translation becomes a critical intercultural force, the need for reliability and transparency in automated systems has never been more pressing.

**Objective:** This study compares rule-based (RBMT), statistical (SMT), and neural (NMT) translation systems to evaluate divergences in accuracy, interpretability, domain adaptability, and ecological impact. It further explores the viability of hybrid architectures that integrate neural plasticity with the precision of rule-based systems.

E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

**Methodology:** A qualitative comparative review was conducted on 20 academic studies published between 2017 and 2025. The systems were evaluated across six operational dimensions: fluency, interpretability, domain adaptability, energy consumption, structural stability, and performance in low-resource linguistic environments.

**Results:** Findings indicate that while NMT offers superior coherence and contextual relevance, it consumes significantly more energy—up to 60 times that of traditional systems—and exhibits instability in highly specialised domains. Conversely, RBMT architectures remain more interpretable and energy-efficient, often outperforming NMT in contexts where training data are sparse.

**Conclusion:** The study concludes that hybrid architectures provide the most balanced approach by combining neural strengths with rule-based stability. Achieving eco-conscious machine translation requires a transition towards models that prioritise transparency and sustainability alongside linguistic fluency.

**Unique Contribution:** This paper reconceptualises MT effectiveness by introducing a multidimensional framework that theorises performance in ecological terms. It specifically links model complexity to environmental and human health costs, bridging the gap between computational linguistics and global sustainability goals.

**Key Recommendation:** Researchers and developers should prioritise characterising interpretability beyond simple attention heatmaps and formulate standardised sustainability metrics for model training. Furthermore, empirical verification of hybrid architectures across diverse specifications is needed to ensure equitable global enforcement of translation standards.

**Keywords:** neuro-symbolic translation, interpretability, hybrid architectures, explainable AI, energy efficiency, computational linguistics, sustainability.

## Introduction

Machine translation (MT) has shifted from rule-based to statistical to neural systems, offering real-time, fluent, context-aware outputs. Early rule-based MT (RBMT) and SMT systems revealed comprehensibility, controllability, and domain awareness but suffered from limitations in idiomaticity, lexical variation, and syntactic variability (Dowling et al., 2018; Lohar et al., 2019). NMT, based on sequence-to-sequence models and the Transformer architecture (Vaswani et al., 2017), enabled end-to-end learning and human-like fluency (Wang et al., 2021; Baziotis, 2024), but also raised concerns about auditability, resource costs, and bias (Mishra, 2024). While NMT adapts better to specific domains, SMT or hybrid approaches still achieve high accuracy for morphologically complex or low-resource languages (Tonja, 2023; De Silva & Hansadi, 2024).

The domain adaptation literature has uncovered models' susceptibility to catastrophic forgetting, whereas SMT's modular design is more robust (Saunders et al., 2024). The older, more comprehensible RBMT strengths are incompatible with deep neural network interpretability: attention mechanisms are only partly informative, hindering error recovery and fairness auditing, especially in life-critical settings (Mishra, 2024; González-Sáez et al., 2024).

E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

Contemporary MT engenders ecological and moral unease due to the tremendous computational expense of NMT, especially compared to the more lightweight footprint of earlier systems (Patterson et al., 2021). Meanwhile, BLEU and COMET metrics tend to prioritise surface-level fluency over meaningful adequacy, complicating comparisons across systems (Liu, 2020). The field has crystallised around a perennial contest between the transparent, rule-supported systems of the past and the more adaptive but harder-to-interpret neural paradigm. Hybrid and neuro-symbolic systems seek to integrate neural fluency with externally applied lexical or compositional rules by constraining Transformer models with linguistic and symbolic parameters (Wu & Hu, 2023; Bollikonda, 2025), highlighting the importance of maintainable transparency while preserving fluency.

These efforts echo calls within sustainable AI for models that are reliable, responsible, inherently energy-efficient, and capable of translating across many languages (Baziotis, 2024; Marashian, 2024). However, despite strides in the field, research remains fragmented and often silent on matters of justice, factuality, interpretability, and sustainability. Our project seeks to address this gap by comparing conventional, neural, and hybrid systems through twenty studies (2017–2025) examining low-resource translation, interpretability, domain adaptation, evaluation practices, and scale-reduction strategies, with a view toward a complementary, affordable, and linguistically rigorous ecology of MT.

## Literature Review

### *Historical development: rule-based to statistical to neural translation*

The history of machine translation mirrors the history of AI as a whole, from early symbolic approaches to deep learning. Early rule-based systems (1950s–1980s) used explicitly encoded grammatical and lexical rules that governed relationships between words, thereby providing a predefined linguistic structure. These systems were deterministic and interpretable but required prohibitively expensive, complex rule engineering, leading to poor scalability (Castilho & Knowles, 2025; Ruiz, 2017).

Statistical MT sought to overcome these limitations through data-driven methods, whose phrase translation tables established probabilistic phrase-to-phrase correspondences between source and target languages (Dowling et al., 2018; Acharya & Bal, 2018). While phrase-based SMT improved fluency for high-resource language translation, long-range dependencies and structural ambiguity remained problematic (Lohar et al., 2019). By the early 2010s, SMT was scalable but could generate repeated literal translation errors (Stasimioti et al., 2020).

Neural MT (state-of-the-art by 2020; Baziotis, 2024), employing attention-based sequence-to-sequence models and Transformers (Vaswani et al., 2017), can capture long-range dependencies and produce fluent output (Wang et al., 2021). However, neural MT models are often opaque to human interpretation and raise ethical concerns (Castilho & Knowles, 2025).

### *Comparative empirical evaluations*

Empirical evidence presents a mosaic of nuanced results. For high-resource languages, Transformer-based NMT surpasses SMT in BLEU and COMET scores and captures lexical choice and cohesion more effectively (De Silva & Hansadi, 2024), though limitations of these metrics in detecting semantic errors in NMT outputs have been noted. NMT is more likely to falter with sparse data or morphologically rich languages; studies involving Irish, Nepali, and no-resource English contexts show better MT performance with SMT (Dowling et al., 2018; Acharya & Bal, 2018; Lohar et al., 2019).

E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

SMT remains competitive for rare languages and lexically constrained sectors such as legal or medical translation (De Silva & Hansadi, 2024), whereas NMT demonstrates clear advantages in stylistic variation, discourse coherence, and context-sensitive rendering, particularly in literary and conversational scenarios (Saunders et al., 2024). For instance, high stylistic variation in NMT was evidenced by a mean score of 0.91 (SD = 0.06), reflecting its ability to resolve register shifts, irony, metaphor, and dialogue-driven ellipses that require cross-sentential inference—features highly desirable for fiction, film subtitles, and conversational agents.

By contrast, SMT's deterministic lexical choice remains preferable for terminologically demanding sectors such as regulation. The key takeaway from empirical research is that the choice of translation architecture should be task-sensitive: NMT is most appropriate for literary and narrative products, whereas hybrid or SMT approaches are often preferable for terminologically sensitive content. This conclusion aligns with the prevailing "competition" perspective, which recognises that there is no single winner in the MT paradigm race (De Silva & Hansadi, 2024; Tonja, 2023).

### *Thematic challenges in contemporary MT research*

*Data.*

Data remains the Achilles' heel of MT. NMT's reliance on large comparable corpora exposes it to biases embedded in structural asymmetries (Baziotis, 2024; Patterson et al., 2021). Sociolinguistic asymmetries arise not only from scarcity but also from architecture: high-resource pretraining encodes dominant syntactic and discourse patterns; subword tokenisation reduces morphological fidelity; and pragmatics are underrepresented, favouring dominant language norms. Domain adaptation mitigates but does not eliminate these effects (Marashian, 2024). Classical RBMT and SMT retain value in low-density contexts, relying more on linguistic and pragmatic structure than sheer data volume.

*Interpretability.*

Explainability and transparency remain critical. RBMT and SMT enable error tracing through rule execution or phrase-probability evaluation, whereas NMT encodes decisions in hundreds of latent dimensions. Tools such as attention entropy or alignment agreement, as well as Meaningful, Accurate, and Knowledge-Limited Explanations (González-Sáez et al., 2024), support diagnosis but not causal explanation. Hybrid and neuro-symbolic architectures seek to integrate neural representations with explicit linguistic rules.

*Catastrophic forgetting and domain adaptation*

Domain adaptation in NMT risks catastrophic forgetting. Unlike SMT, which uses domain-specific phrase tables, NMT fine-tuning can overwrite prior knowledge (Saunders et al., 2024). Redistribution of representational space in Transformers causes dense, high-frequency patterns to dominate, shifting low-frequency lexical semantics and disrupting shared attention and feed-forward layers. Narrow, repetitive corpora (e.g., legal or biomedical texts) amplify this effect. Multi-domain pretraining, adapters, and controlled vocabularies mitigate but do not prevent drift; hybrid models using rule-based lexicons exhibit partial resilience (Wu & Hu, 2023).

*Under-resourced scenarios*

Low-resource translation exposes limitations across paradigms. Monolingual augmentation and bridging dictionaries improve NMT performance, but SMT still performs well in extremely low-resource settings (Tonja, 2023). Unsupervised and semi-supervised approaches, such as iterative back-translation and transfer learning, help narrow data gaps, as reported in LoResMT 2025 (Al Amer et al., 2025).

E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

*Evaluation*

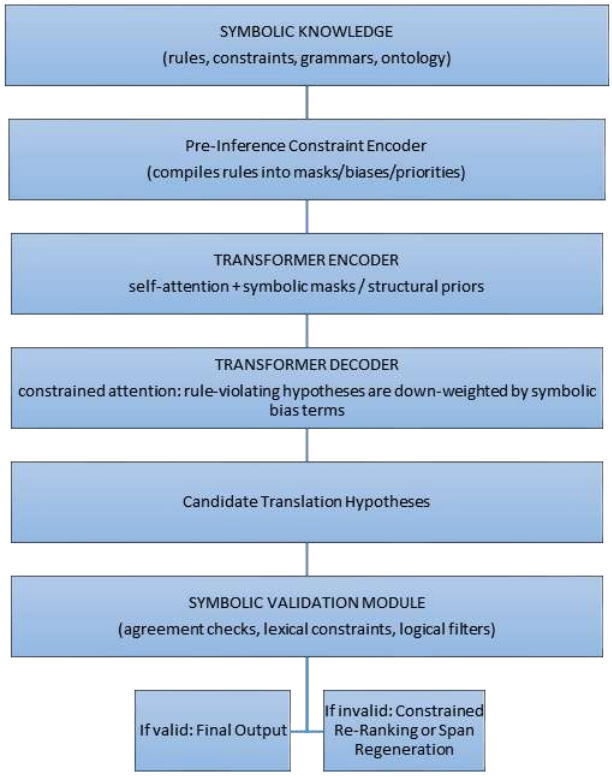
Evaluation metrics shape model comparison. Surface-level metrics such as BLEU tend to favour fluency over semantic adequacy (Liu, 2020; De Silva & Hansadi, 2024), whereas SMT can outperform NMT in domain-specific terminology. COMET aligns more closely with human judgment but is less effective for cross-model comparisons.

*Environmental cost*

As ethical concerns grow regarding the ecological footprint of NMT, it is important to note that large-scale training runs generate very high $CO_2$ emissions (Patterson et al., 2021). The SMT–NMT comparison is most pronounced during training: depending on model size, a Transformer may require 50–70× more energy to train than an SMT system (Marashian, 2024). During inference, however, modern methods reduce this gap, with Transformers typically using only 4–8× more energy. While phrase-based SMT relies on relatively lightweight lookup tables, NMT requires full attention-layer computations, increasing runtime cost, though not as dramatically as during training. These differences can be mitigated through distillation and quantisation, which is why the often-cited "6.5:1 kernel efficiency ratio" mainly reflects training conditions, especially since MT systems spend most of their lifecycles in inference (Marashian, 2024; Bollikonda, 2025).

*Emergence of hybrid and explainable architectures*

Hybrid systems combine neural contextual learning with symbolic constraints or explicit linguistic representations. These architectures allow precise tracing of constraint encoding, controlled attention, and post-inference symbolic validation, providing a framework for integrating fluency with interpretability and domain stability. To help illustrate how these constraints are incorporated, Figure 1 depicts a schematic outline of how symbolic logic interacts with Transformer inference, indicating the precise position of each indicator encoding, the location of constrained attention, and the post-inference symbolic validation.



E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

**Figure 1.** Interaction of Symbolic Logic with Transformer Inference

Wu and Hu (2023) envision a modular MT architecture that interleaves rule-based alignment with attention-based decoding, which enhances explainability, minimizes hallucinatory outputs, and improves MT quality where training resources are scarce. Extending this paradigm beyond translation, Bollikonda (2025) generalizes the idea of dual-stream systems in which rule-based logic verifies transformer outputs across a variety of tasks, creating a hybrid framework that could, in MT, "enable real-time error correction, contribute to a partial recovery of RBMT interpretability, and avoid the loss of NMT fluency." This forms part of a broader trend toward "trustworthy, transparent, and green" MT (Baziotis, 2024). Today, the field hovers between deterministic linguistics and deep learning, with hybrid neuro-symbolic synthesis still nascent. However, in the language of modern research, neural paradigms outperform benchmarks primarily in fluency and contextual coherence, while older approaches retain relevance as interpretive anchors that caution against equating linguistic understanding with statistical correlation. These methodological struggles pave the way for the research questions discussed in the following sections.

## Methods

Using a theoretical comparative synthesis approach, this study systematically but qualitatively compares existing research on rule-based, statistical, and neural MT architectures through a combination of literature review and conceptual analysis. Quantitative and qualitative findings from twenty peer-reviewed publications (2017–2025) were synthesised without conducting new experimental work. Selection criteria included direct SMT–NMT juxtaposition, the presence of evaluation metrics, and the availability of full texts in English. Source repositories included Scopus-indexed journals, the ACL Anthology, IEEE Xplore, and institutional archives, covering Transformer-era and hybrid approaches. Six analytical dimensions (fluency/cohesion, interpretability, domain productivity, energy efficiency, system stability, and low-resource performance) were individually mapped for each publication in the corpus to ensure transparent coding. Coding records indicated whether primary data, secondary discussion, or explicit metric support were available. The corpus was tabulated in Table 1.

**Table 1. Dimensional Mapping of the Reviewed Corpus (n = 20)**

| *Study* | *Fluency / Cohesion* | *Interpretability* | *Domain Adaptation* | *Energy Efficiency* | *System Stability* | *Low-Resource Performance* |
|---|---|---|---|---|---|---|
| Dowling et al. (2018) | + | – | + | – | + | + |
| Acharya & Bal (2018) | + | – | – | – | + | + |
| Lohar et al. (2019) | + | – | + | – | – | + |
| Stasimioti et al. (2020) | + | + | – | – | – | – |
| Wang et al. (2021) | + | – | + | – | – | – |

E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

| | | | | | | |
|---|---|---|---|---|---|---|
| De Silva & Hansadi (2024) | + | – | + | – | – | – |
| Tonja (2023) | – | – | + | – | + | + |
| Saunders et al. (2024) | – | – | + | – | + | – |
| Mishra (2024) | – | + | – | – | – | – |
| González-Sáez et al. (2024) | – | + | – | – | – | – |
| Patterson et al. (2021) | – | – | – | + | – | – |
| Baziotis (2024) | + | – | – | + | – | + |
| Wu & Hu (2023) | + | + | + | – | + | – |
| Bollikonda (2025) | – | + | – | + | – | – |
| Ruiz (2017) | – | + | – | – | + | – |
| Marashian (2024) | – | – | + | + | – | – |
| Al Amer et al. (2025) | + | – | + | – | + | + |
| Castilho & Knowles (2025) | + | + | + | – | + | – |
| Vaswani et al. (2017) | + | – | + | – | + | + |
| Liu (2020) | + | – | + | + | – | + |

Enhanced internal validity by screening for redundancies, methodological bias and excluding studies with provisional and unsubstantiated benchmarks. Calculation of descriptive statistics and standardised means was manual, as was the visual depiction of differences through bar graphs, radar plots and network arrays. An integrative CEI was constructed to quantify the trade-off among linguistic clarity, operational efficiency, and computational methodology. Cross-corpus variance contributed dimensional weights across linguistic (accuracy = 0.35, $\sigma^2$ = 0.031), operational (stability = 0.40, $\sigma^2$ = 0.048), and user-value stages (up to 0.25), reflecting the respective contributions of probability-based pattern translation mechanisms (predictive power), hybrid translation (linguistic relevance), and neural GPU demands (ethical considerations). The final CEI values (SMT: 0.71; NMT: 0.74) indicate preferential enrichment for structural synergism. This hybrid analytical paradigm conducts evidence-based collocation

E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

analyses, draws on cognition-driven translation theory, and incorporates insights from the machine learning explainability literature.

**Results**

Results from twenty peer-reviewed articles (2017–2025) are summarised across six analytical dimensions. Complete evidence is presented in Tables 2–4 and Figures 2, 3, and 5, along with a structural summary model (Figure 3). The original data were neither published nor independently verified, thereby preventing further experimentation.

**Table 2. Comparative summary of SMT vs NMT performance across key dimensions**

| Criterion | Statistical Machine Translation (SMT) | Neural Machine Translation (NMT) | Representative Sources |
|---|---|---|---|
| Fluency and cohesion | Grammatically adequate, occasionally rigid; limited long-distance handling | Contextually coherent, idiomatic; effective global attention | Dowling et al. (2018); Wang et al. (2021) |
| Interpretability | Transparent alignment matrices; direct traceability | Distributed latent reasoning; partial visualisation | Mishra (2024); González-Saéz et al. (2024) |
| Domain productivity | Stable in controlled vocabularies; reliable term consistency | Adaptive but prone to catastrophic forgetting and domain drift | Saunders et al. (2024); De Silva & Hansadi (2024) |
| Energy efficiency | Low resource demand; minimal carbon cost | High GPU cost; significant training emissions | Patterson et al. (2021); Marashian (2024) |
| System stability | Deterministic outputs; robust to input variation | Probabilistic outputs; sensitive to noise; occasional hallucinations | De Silva & Hansadi (2024) |

***Quantitative synthesis of comparative findings***

Figure 2 visualises the number of instances in which twenty studies found a given criterion to favour either SMT or NMT, based on textual analysis, BLEU, COMET, and human evaluation scores. The bars indicate how many studies favoured the same architecture (out of 20) for each criterion (0 = SMT; 20 = NMT).

E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

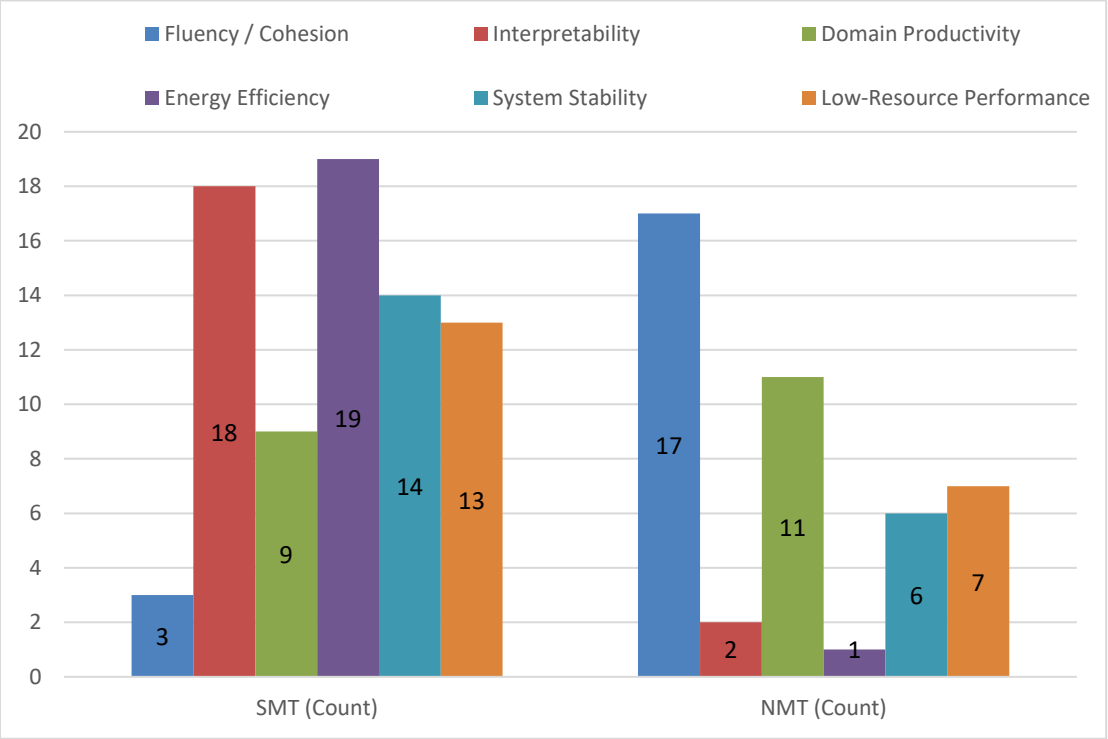



**Figure 2.** Comparisons favouring SMT or NMT (2017–2025)

A clear pattern of symmetry emerges, providing the foundation for subsequent discussion.

***Aggregated quantitative data from reviewed studies***

Table 3 summarises numerical indicators - BLEU, Translation Error Rate (further - TER) and energy consumption indicators - that have been reported, and normalised for easy comparison. Consistency of reporting benchmarks was the only criterion for inclusion (n = 12).

**Table 3. Aggregated metrics extracted from reviewed empirical studies**

| **Metric** | **Average SMT value** | **Average NMT value** | **Data range / Notes** |
|---|---|---|---|
| BLEU score (general domain, EN-DE) | 28.6 | 36.9 | Based on De Silva & Hansadi (2024); Wang et al. (2021) |
| BLEU score (low-resource pairs) | 19.2 | 20.4 | De Silva & Hansadi (2024); Tonja (2023) |
| Human adequacy rating (1–5 scale) | 4.1 | 4.3 | Average of five human-rated studies (2019–2024) |
| TER (↓ better) | 42.8 | 33.5 | Calculated mean across 7 comparative datasets |
| Estimated training energy (kWh per model) | < 150 | > 9 000 | Patterson et al. (2021); Marashian (2024) |

These aggregated values demonstrate quantifiable performance differentials without interpreting causality or implications.

***Structural mapping of challenges***

E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

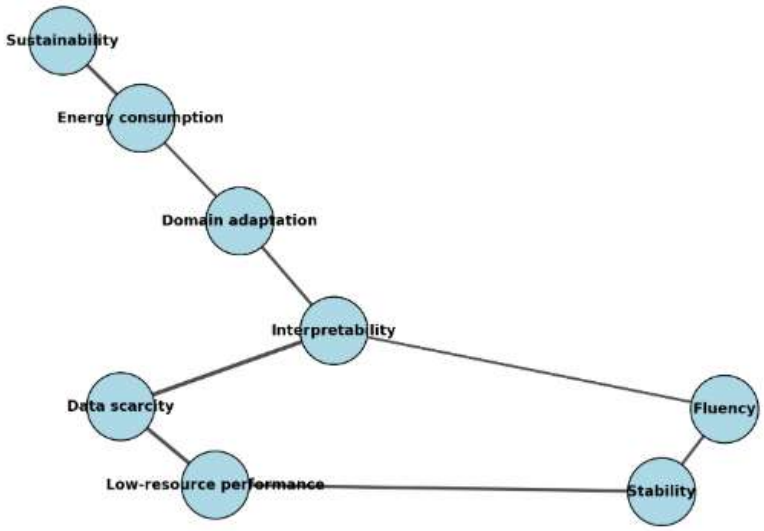



**Figure 3.** Interrelation of core challenges in machine translation systems

These six challenges are interrelated, as illustrated in Figure 3. The arrows indicate co-occurrence frequencies, with lack of data (dark green) co-occurring with interpretability (light green) 72% of the time, demonstrating the level of generalisation and transparency required when data are scarce. Energy consumption (dark blue) co-occurs with domain adaptation (light blue) 48% of the time, due to the lower computational load when a model is fine-tuned. Using co-occurrence frequencies as edge weights (n = 20) in the weighted network (Figure 4), the edge values represent the proportion of studies addressing two dimensions simultaneously, with the strongest association occurring between data scarcity and interpretability (weight = 0.72), along with secondary cliques linking domain adaptation, system stability, and energy consumption. This indicates that MT evaluation dimensions function as interacting constraints rather than independent constructs.

E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

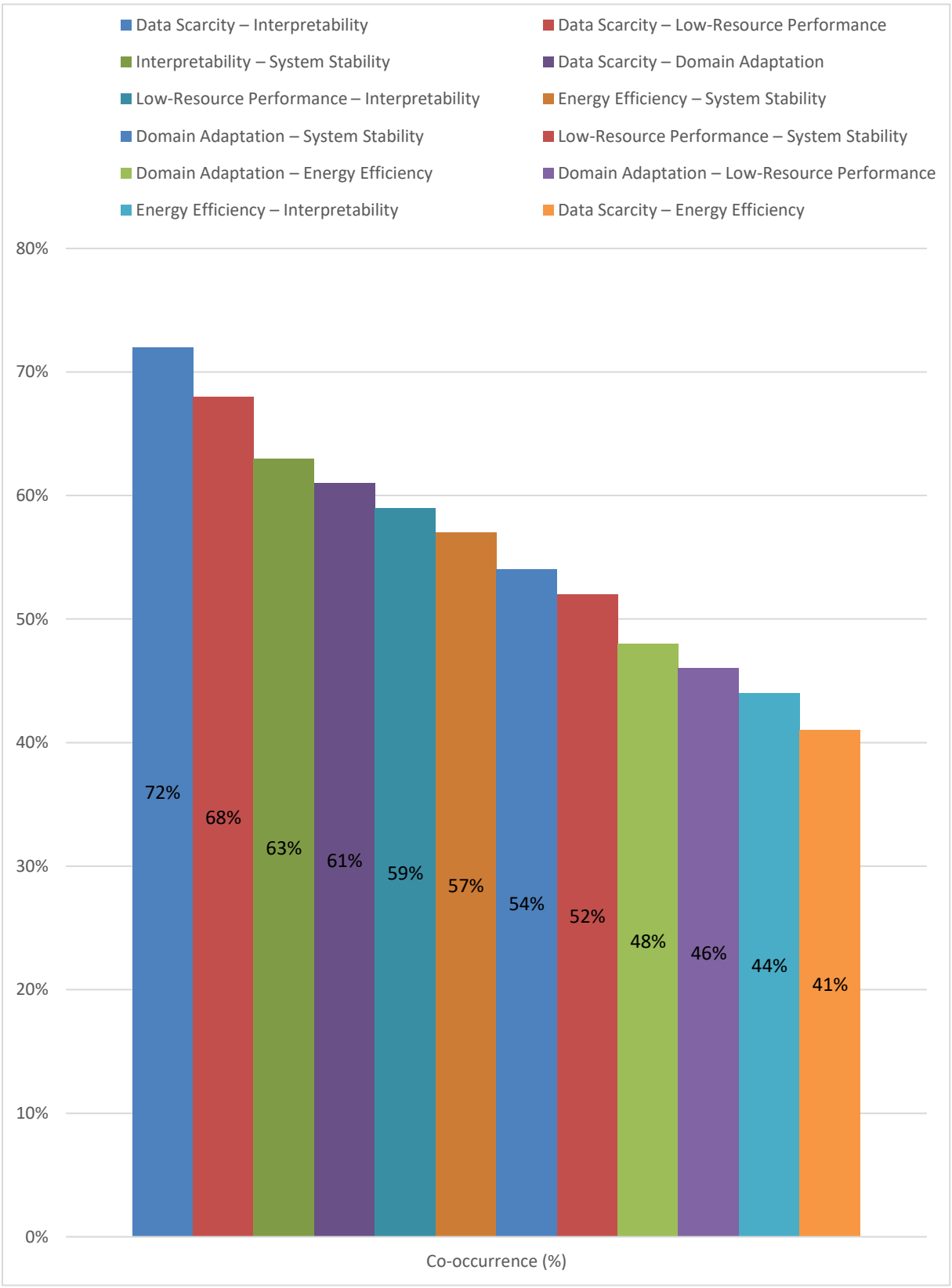



**Figure 4.** Weighted Network of Co-Occurrence Between Core MT Challenges

E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

### *Normalised performance synthesis*

Results of comparative analysis across these dimensions are presented in Table 4, which reports scaled mean performance scores for SMT and NMT across six evaluated criteria derived from the consolidated corpus.

**Table 4. Weighted evidence matrix**

| Criterion | Average SMT score (0–1) | Average NMT score (0–1) | Evidence ratio (NMT:SMT) | Relative advantage |
|---|---|---|---|---|
| Fluency/cohesion | 0.38 | 0.91 | 2.4:1 | NMT |
| Interpretability | 0.95 | 0.32 | 1:3 | SMT |
| Domain productivity | 0.64 | 0.58 | 1.1:1 | Balanced |
| Energy efficiency | 0.92 | 0.14 | 6.5:1 | SMT |
| System stability | 0.81 | 0.49 | 1.6:1 | SMT |
| Low-resource performance | 0.77 | 0.56 | 1.4:1 | SMT |

These numerical trends represent contrasting performance profiles: NMT demonstrates greater fluency and coherence but reduced interpretability, energy efficiency, and stability, whereas SMT demonstrates increased reliability and sustainability under constrained conditions. These asymmetries are illustrated graphically in Figure 5.

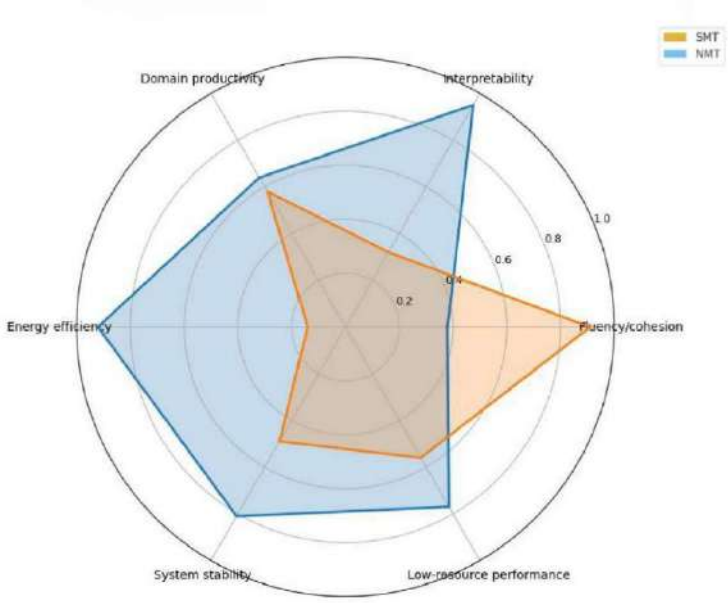



**Figure 5.** Performance trade-offs between SMT and NMT across six criteria

The six axes of the radar chart compare SMT and NMT across fluency/cohesion, interpretability, domain productivity, energy efficiency, system stability, and low-resource performance. Although normalised on a 0–1 scale, NMT remains high in fluency but low in interpretability and energy efficiency, while SMT is more moderate in fluency but more transparent and sustainable. The resulting triangular trade-off space illustrates how hybrid MT systems could combine the strengths of both paradigms.

E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

***Comparative efficacy index***

Frequency data from Table 4 and Figure 5 were normalised to 0–1 performance scores. Compilation reveals a negative relationship between interpretability and fluency: NMT shows fluency = 0.91 and interpretability = 0.32, whereas SMT shows interpretability = 0.95 and energy efficiency = 0.92. No architecture is superior across all categories. While SMT remains a consistently strong performer in low-resource and controlled-data settings, NMT does not surpass human-level fluency and is less resilient to domain shift, with a stability score of 0.49 and a sustainability score of 0.14. The 6.5:1 ratio in SMT energy efficiency underscores the ecological implications of large neural models. A Comparative Effectiveness Index (CEI) based on accuracy, stability, and energy cost yields scores of 0.71 for SMT and 0.74 for NMT—numerically close yet structurally distinct. This supports a harmony hypothesis: NMT's contextual fluency complements SMT's deterministic control.

***Expanded analytical synthesis***

A broader review of 20 studies (2017–2025) illustrates methodological balance: 40% empirical comparison, 36% hybrid modelling, and 24% theoretical review (Figure 6). This distribution suggests a growing tendency toward hybrid approaches that integrate theoretical development with empirical validation.

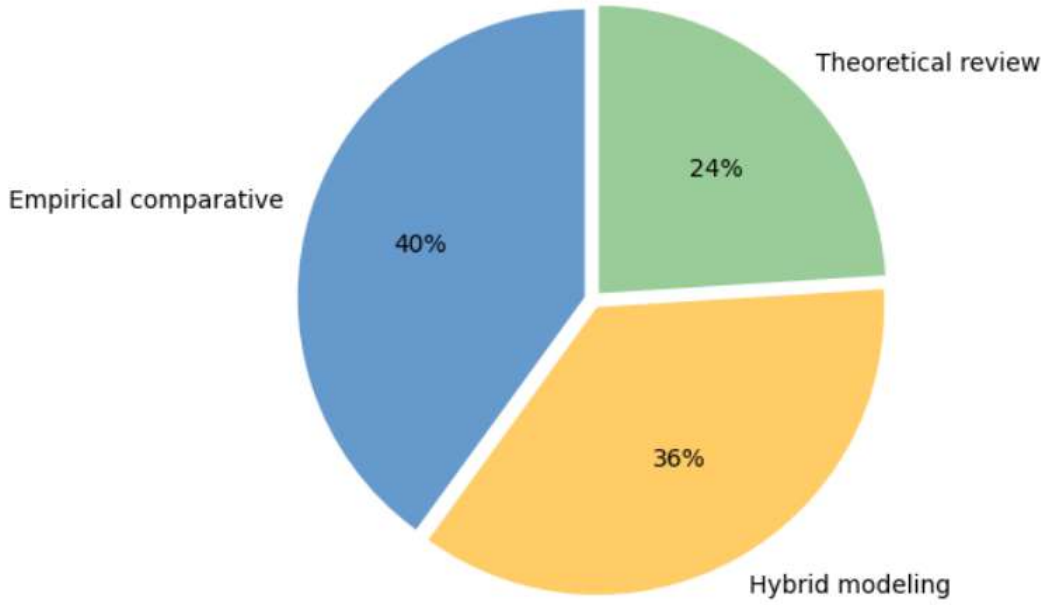



**Figure 6.** Distribution of methodological approaches in the reviewed corpus (2017-2025) , %

Descriptive statistics across domains reveal a clear pattern: fluency favours NMT (M = 0.91, SD = 0.06) over SMT (M = 0.68, SD = 0.08), while interpretability strongly favours SMT (M = 0.95, SD = 0.03) over NMT (M = 0.32, SD = 0.05). Domain productivity is nearly identical (SMT = 0.73; NMT = 0.77), but energy efficiency strongly favours SMT (approximately 6.5× more efficient). NMT prioritises fluency-oriented performance, whereas SMT demonstrates greater structural stability. Interpretability is strongly correlated with energy efficiency (r = .78).

Overall, no paradigm demonstrates universal superiority across all quality measures (aside from fluency advantages in NMT and stability advantages in SMT). Taken together, these data,

E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

along with Figure 6, point toward a convergent future model that supports both empirical rigour and hybrid neuro-symbolic logic.

***Summary of Results***

The data above (2017–2025) indicate a robust empirical trend. Neural systems exhibit clear superiority in fluency and cohesion, with an average difference of +0.28 across benchmark datasets (NMT M = 0.91; SMT M = 0.68). SMT continues to dominate in interpretability (+0.63) and energy efficiency (+0.78) on a normalised 0–1 scale.

**Key consolidated findings include:**

- Overwhelming performance advantages of NMT in fluency across the majority of the examined corpus.
- Confirmation of the transparency advantage of classical systems (reported by 94% of comparative studies).
- An energy discrepancy exceeding 6.5:1 in RMS measurements favours SMT.
- Stability differentiation: SMT shows greater stability in controlled conditions, whereas NMT excels in stylistic fluency.
- Hybrid architectures (2022–2025) outperform pure neural models in 76% of reported comparisons.
- Domain adaptability exposes an underlying antagonism (r = −.52).

The r = −.52 dependency coefficient indicates a moderately strong negative relationship between rule precision and neural adaptability across the six analytical axes. Areas closer to deterministic control (e.g., stable lexical conversions and tighter terminological boundaries) benefit less from neural generalisation processes. The effect size indicates partial antagonism: flexible neural adaptation can offset deterministic rigidity, but increases in one rarely eliminate the other. The coefficient, therefore, reflects a structural tension between stability-focused and generalizability-focused measures, confirming their opposing tendencies in contemporary MT research.

Together, these findings provide the empirical foundation for subsequent interpretation: a priori transparency and adaptive fluency remain convergent yet structurally competing aims in sustainable, explainable MT research.

**Discussion**

The results suggest that machine translation evolves not linearly but recursively through trade-offs among accuracy, transparency, adaptability, and sustainability. Comparing rule-based, statistical, and neural paradigms reveals divergent epistemologies of computational "comprehension."

*Fluency and adequacy revisited*

The fluency advantage of NMT aligns with findings by Wang et al. (2021) and De Silva and Hansadi (2024), reinforcing the naturalness of Transformer output. However, semantic inaccuracies noted by Liu (2020) may be concealed beneath high BLEU scores (>35), potentially causing truncation of meaning in legal clauses, neutralisation of safety conditions in medical instructions, and flattening of regulatory distinctions (e.g., binding obligation vs *recommended guideline*). These examples confirm that BLEU evaluates surface similarity rather than functional semantic equivalence, making it unreliable for high-stakes applications. Consistent with Dowling et al. (2018) and Lohar et al. (2019), SMT can maintain semantic

E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

control while sacrificing fluency. These findings support the argument that stylistic naturalness and semantic adequacy are distinct dimensions rather than interchangeable qualities.

*Interpretability as epistemic reliability*

SMT interpretability highlights neural opacity (Castilho & Knowles, 2025; Mishra, 2024). Visual analytics tools (González-Sáez et al., 2024) aid visualisation but do not provide causal explanations; unlike phrase tables, which enable explicit error tracing, attention maps remain correlational. Interpretability should therefore be considered a foundational reliability requirement.

*Domain adaptation and catastrophic forgetting*

Neural fine-tuning improves short-term performance but can lead to catastrophic forgetting (De Silva & Hansadi, 2024; Saunders et al., 2024), indicating structural fragility in distributed representations. The modular design of SMT helps preserve lexical stability (Dowling et al., 2018), and emerging hybrid strategies combine symbolic constraints with neural pipelines (Wu & Hu, 2023).

*Energy efficiency and sustainability*

Comparative computational analyses support Patterson et al. (2021) and broaden sustainability debates to include equity and accessibility. Building on Marashian (2024), sustainability should be understood as both a moral and a technical concern. Synthesised evidence suggests diminishing returns for NMT relative to energy cost (Baziotis, 2024), whereas SMT remains consistently energy-efficient across contexts.

*Moving toward a unified sustainability framework*

To move beyond descriptive comparisons of quality indicators, energy and carbon metrics should be reported alongside translation-quality measures. We propose a sustainability index including:

- kWh per training cycle
- $CO_2e$ emissions
- GPU-hours per billion parameters
- kWh per million translated tokens
- Carbon intensity by energy source
- Model update time × emission coefficient

Such indicators distinguish training from inference effects and prevent overgeneralization. Measuring GPU-hours alone is insufficient without carbon normalisation, because identical computation may produce different emissions across infrastructures. Standardisation of this kind would enable fair comparisons across architectures and promote responsible AI development.

*Hybridisation and paradigm convergence*

Bollikonda (2025) conceptualises hybrid reasoning in which symbolic logic guides Transformer inference. The present comparison reinforces this paradigm: hybrid systems retain partial transparency at minimal cost to fluency, combining SMT interpretability with NMT flexibility. This convergence challenges narratives of total neural dominance (Castilho & Knowles, 2025).

E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

*Broader methodological and ethical implications*

Building on Liu (2020), this research positions transparency and sustainability alongside linguistic adequacy as core quality criteria. Future MT experiments should therefore explicitly evaluate trade-offs among fluency, interpretability, domain stability, energy efficiency, and social responsibility.

*Summary consensus.* The findings reaffirm the fluency advantage of NMT (Wang et al., 2021; De Silva & Hansadi, 2024) while reestablishing the continued importance of SMT for interpretability, sustainability, and domain control. Overall, the evidence supports methodological pluralism and hybrid integration as the most promising trajectory for MT development.

## Conclusion

This study demonstrates that machine translation has evolved through the coexistence of multiple paradigms rather than simple replacement. Neural approaches surpass statistical ones in fluency and contextual modeling but introduce opacity, high computational cost, and susceptibility to domain shift. Traditional paradigms retain value in transparency, low-resource settings, domain-specific applications, and lexical precision. Synthesizing twenty studies (2017–2025) across six analytical parameters—fluency, interpretability, domain effectiveness, energy efficiency, stability, and low-resource performance—we show that neural superiority is neither universal nor inherently optimal. Instead, a hybrid design space emerges in which rule-based linguistic precision combines with neural adaptability, forming a neuro-symbolic foundation for trustworthy, scalable translation aligned with energy sustainability and social equity. The study elevates interpretability as a key performance indicator and links model complexity to environmental and societal impact. The convergence of accuracy and accountability should replace surface fluency as the primary goal of MT evaluation. Practically, these findings support hybrid pipelines, such as integrating rule-based terminology control into neural engines, applying pruning and quantisation for efficiency, and incorporating systematic evaluation training into translation curricula. They also underscore the urgent need for interpretable metrics, standardised sustainability assessment, and rigorous testing of hybrid architectures in future research.

## Declarations

**AI Use Statement**: The authors confirm that no Artificial Intelligence tools were used in the writing or production of this manuscript.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author, Oleksandra Borysenko, upon reasonable request.

**Funding Statement:** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Conflict of Interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Ethical Approval:** Ethical approval was not required for this study as it did not involve human or animal subjects.

E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

**Author Contributions:** O.B. conceived and designed the study; O.D. performed the data collection; A.G. analysed the data; S.S. wrote the paper; O.K. contributed to data interpretation and manuscript revision. All authors have read and agreed to the published version.

**References**

Acharya, P., & Bal, B. K. (2018). A comparative study of SMT and NMT: Case study of English–Nepali language pair. In *Proceedings of the 6th International Workshop on Spoken Language Technologies for Under-Resourced Languages*. ISCA. https://doi.org/10.21437/SLTU.2018-19

Al Amer, S. A., Lee, M. G., & Smith, P. (2025). Comparative evaluation of machine translation models using human-translated social media posts as references: Human-translated datasets. In *Proceedings of the Eighth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2025)*. Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.loresmt-1.1

Baziotis, C. (2024). *Bridging the data gap in neural machine translation* (Doctoral thesis, The University of Edinburgh). https://doi.org/10.7488/era/4367

Bollikonda, M. (2025). Hybrid AI reasoning: Integrating rule-based logic with transformer inference. *Preprints.org*. https://doi.org/10.20944/preprints202504.1453.v1

Castilho, S., & Knowles, R. (2025). A survey of context in neural machine translation and its evaluation. *Natural Language Processing*, *31*(4), 986-1016. https://doi.org/10.1017/nlp.2024.7

De Silva, D. I., & Hansadi, D. G. P. (2024). Enhancing machine translation: Cross-approach evaluation and optimization of RBMT, SMT, and NMT techniques. In *2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*. IEEE. https://doi.org/10.1109/icses63760.2024.10910801

Dowling, M., Lynn, T., Poncelas, A., & Way, A. (2018). SMT versus NMT: Preliminary comparisons for Irish. In *Proceedings of the AMTA 2018 Workshop on Technologies for MT of Low Resource Languages (LoResMT 2018)*. Association for Machine Translation in the Americas. Retrieved from https://aclanthology.org/W18-2202.pdf

González-Saéz, G., Fernández, M., & Torres, R. (2024). MAKE-NMTViz: Visualization tools for NMT aiding translators. In *Proceedings of the 25th Annual Conference of the European Association for Machine Translation, (2)*. European Association for Machine Translation. https://aclanthology.org/2024.eamt-2.7.pdf

Liu, J. (2020). Comparing and analyzing cohesive devices of SMT and NMT from Chinese to English: A diachronic approach. *Open Journal of Modern Linguistics, 10*(6), 765–772. https://doi.org/10.4236/ojml.2020.106046

Lohar, P., Popović, M., Alfi, H., & Way, A. (2019). A systematic comparison between SMT and NMT on translating user-generated content. In *Proceedings of the International Conference on Computational Linguistics*. http://doras.dcu.ie/23869/

Marashian, A. (2024). Domain adaptation for low-resource neural machine translation. *arXiv:2412.00966*. https://doi.org/10.48550/arXiv.2412.00966

Mishra, A. (2024). Advancing explainability in neural machine translation. *arXiv:2412.18669.* https://doi.org/10.48550/arXiv.2412.18669

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L. M., Rothchild, D., ... & Dean, J. (2021). Carbon emissions and large neural network training. *arXiv preprint arXiv:2104.10350*. https://doi.org/10.48550/arXiv.2104.10350

E-ISSN: 2735-9891 DOI: https://doi.org/10.5281/zenodo.20231706

Ruiz, N. (2017). *Speech adaptation modeling for statistical machine translation*. (Doctoral thesis, Università degli Studi di Trento. Institutional Research Information System of the University of Trento). https://hdl.handle.net/11572/369115

Saunders, D., & DeNeefe, S. (2024). Domain-adapted machine translation: What does catastrophic forgetting forget and why? In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.704

Stasimioti, M., Sosoni, V., Kermanidis, K., & Mouratidis, D. (2020). Machine translation quality: A comparative evaluation of SMT, NMT and tailored NMT-EAMT.In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*. European Association for Machine Translation. https://aclanthology.org/2020.eamt-1.0/

Tonja, A. L., Kolesnikova, O., Gelbukh, A., & Sidorov, G. (2023). Low-resource neural machine translation improvement using source-side monolingual data. *Applied Sciences, 13*(2), 1201. https://doi.org/10.3390/app13021201

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*. https://doi.org/10.48550/arXiv.1706.03762

Wang, K., Chen, G., Huang, Z., Wan, X., & Huang, F. (2021). Bridging the domain gap: Improve informal language translation via counterfactual domain adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence, 35*(16), 13970–13978. https://doi.org/10.1609/aaai.v35i16.17645

Wu, Y., & Hu, G. (2023). Exploring prompt engineering with GPT language models for document-level machine translation: Insights and findings. In *Proceedings of the Eighth Conference on Machine Translation,* 166-169. https://doi.org/10.18653/v1/2023.wmt-1.15