Machine Translation

Breaking Down Language Barriers with AI

ITAI 2373: Module 11

Learning Outcomes

- Trace the evolution from rule-based to neural translation.
- Explain encoder–decoder & Transformer architectures.
- Explore real-world applications and business impact
- Examine ethical considerations in cross-cultural communication.

The Translation Challenge

- Languages have different structures and rule
- Cultural context shapes meaning
- Ambiguity exists in all languages
- Idioms and expressions don't translate literally
- One word can have multiple meanings

[1]

Evolution of Machine Translation

1950s

Rule-based systems emerge

1990s

Statistical methods dominate

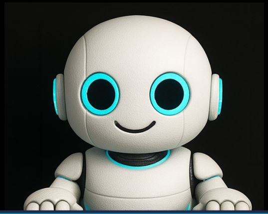
2010s

Neural networks revolutionize the field

2020s

Transformer models achieve near-human quality

MT Paradigms

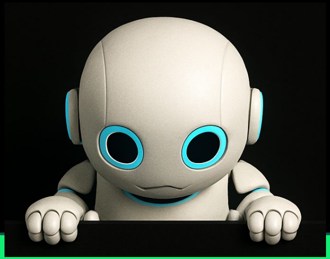


Rule-based

- Uses linguistic rules and dictionaries
- Requires expert knowledge for each language pair
- Predictable and consistent output
- Struggles with ambiguity and exceptions

[2]

MT Paradigms



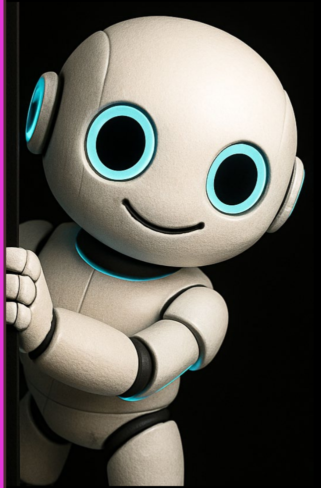
Statistical

- Learns from large collections of translated texts
- Uses probability to choose best translations
- Requires parallel corpora for training
- Dominated the field from 1990s to 2010s

MT Paradigms

Neural

- Uses deep learning to understand context
- Produces more fluent and natural translations
- Can handle longer sentences and complex structures
- Requires large amounts of training data



Neural MT Architecture

Encoder

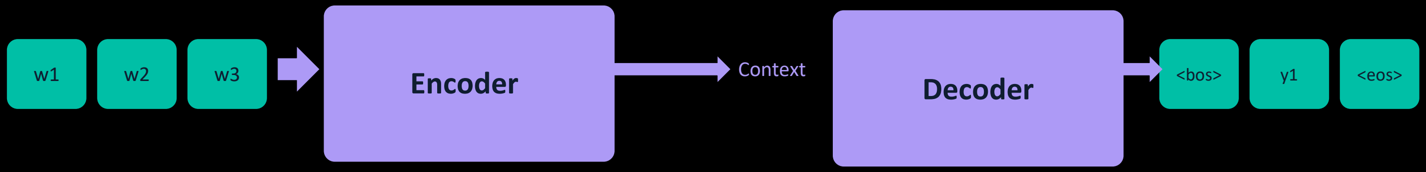
- Compress source tokens input sequence to build a hidden representation.

Decoder

- Generates output tokens one by one conditioned on the context vector.

Special tokens

- <bos> marks start, <eos> marks end of sequence.



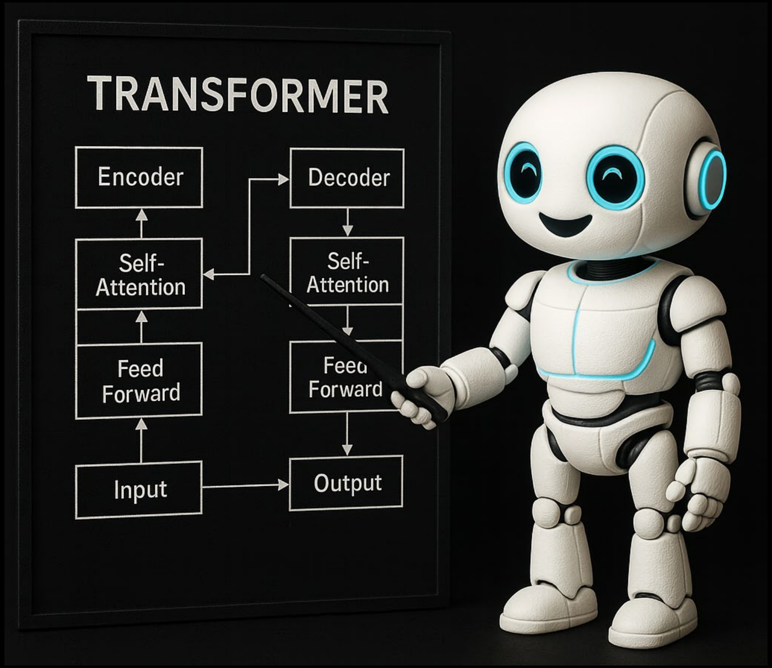
[3]

Google Cloud

Encoder - Decoder Architecture

The Transformer Revolution

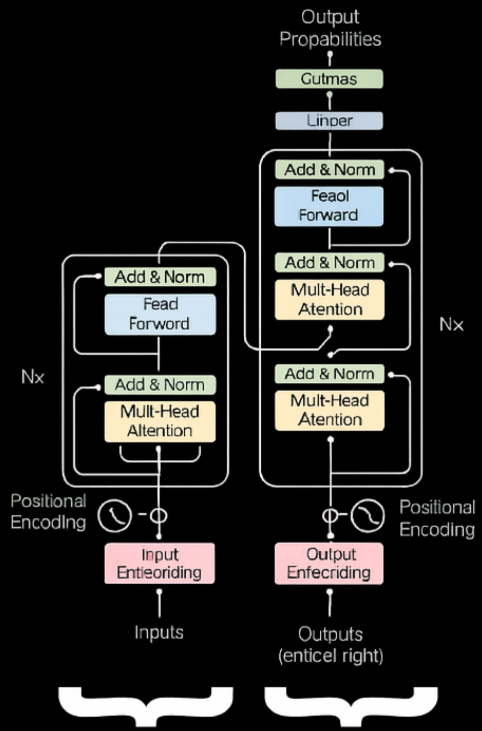
- **Attention mechanisms** focus on relevant parts of text
- **Parallel processing** enables faster training
- Better handling of **long-range** dependencies
- **Foundation** for modern translation systems
- Multi-head attention: multiple perspectives of context
- **Transformers:** replace recurrence with attention and positional encodings



Google Cloud

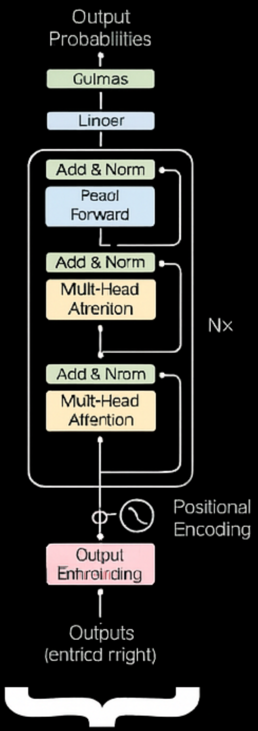
Attention Mechanism: Overview

Transformer



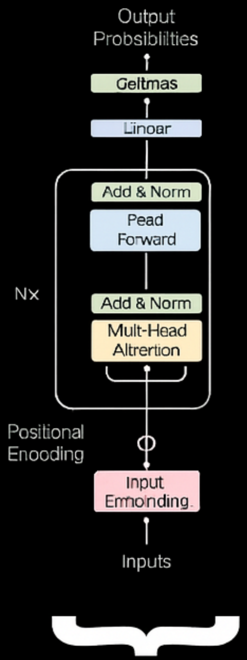
Encoder **Decoder**

GPT*

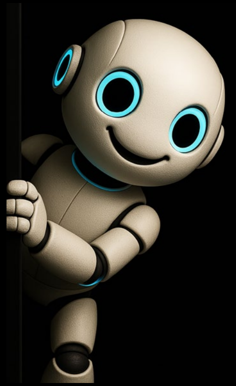


Encoder-only

BERT*



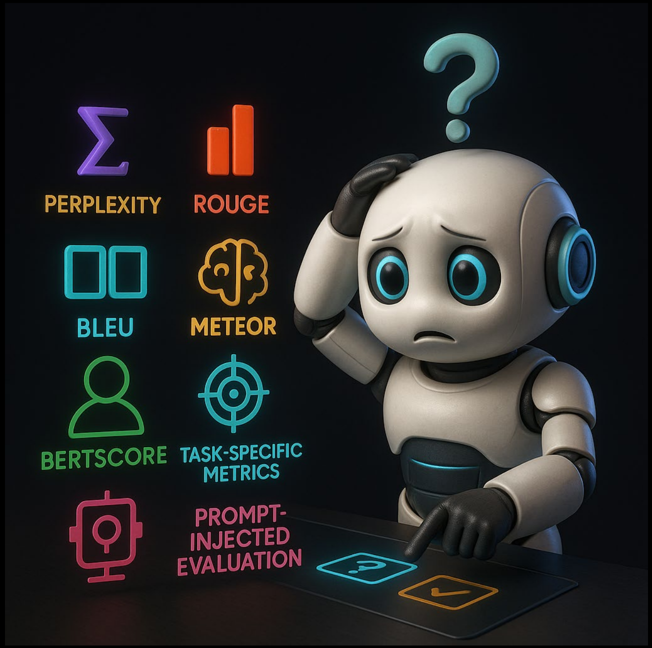
Encoder-only



*Illustrative example, exact model architecture may vary slightly

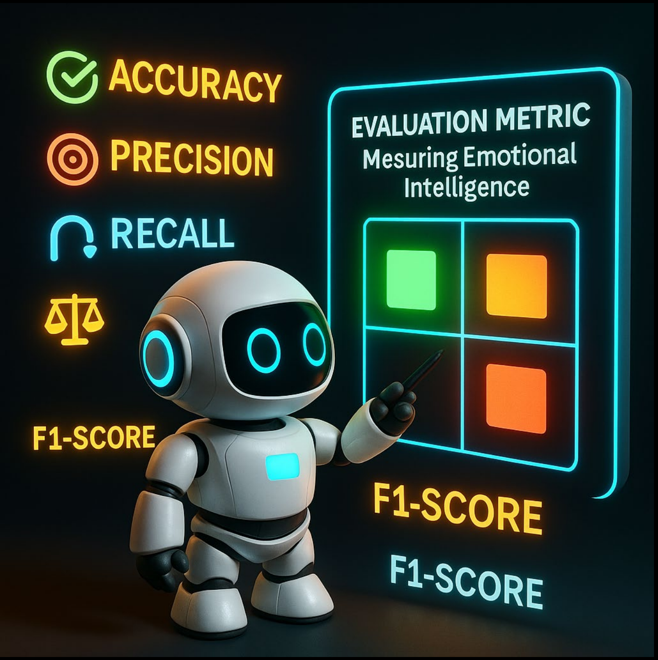
MT Evaluation Challenges

- Many correct translations for a single sentence
- Automatic metrics are proxies for quality
- Human judgment required for fluency & adequacy



[7]

Evaluation Metrics
Three Evaluation Paradigms



Deep Learning (DL) Metrics

- Foundation metrics for neural network training

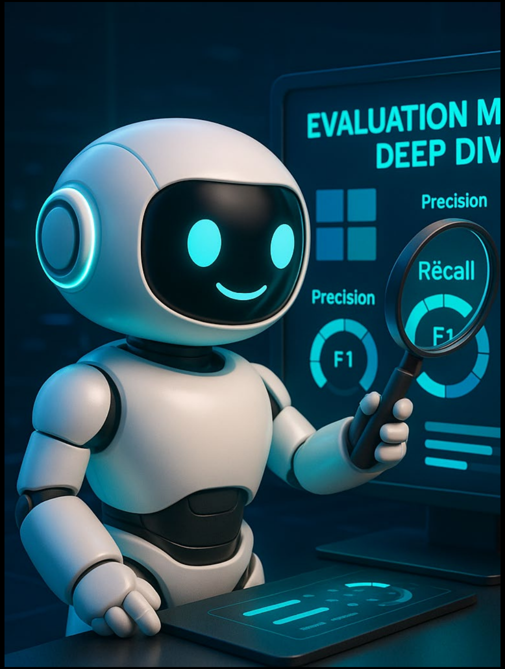
Large Language Model (LLM) Metrics

- Specialized metrics for text generation quality

Machine Translation (MT) Metrics

- Purpose-built metrics for cross-language preservation

DL Metrics - The Foundation



Training Metrics

- Loss Functions: Cross-entropy, MSE, Binary cross-entropy
- Gradient Norms: Learning signal strength monitoring
- Learning Rate: Convergence optimization tracking

Performance Metrics

- Accuracy: Basic correctness measurement
- AUC-ROC: Threshold-independent classification
- Perplexity: Language modeling uncertainty

Primary Use: Model training, convergence monitoring, basic performance

LLM Metrics - Generation Quality

Semantic Understanding

- BERTScore: Contextual embedding similarity
- Semantic Similarity: Meaning-based comparison

Human-Centric Evaluation

- Helpfulness: User utility assessment
- Harmlessness: Safety and bias evaluation
- Honesty: Factual accuracy verification

Task-Specific

- Pass@k: Code generation success rate
- Exact Match: Factual QA correctness

Primary Use: Open-ended generation, subjective quality, user satisfaction

MT Metrics - Cross-Language Fidelity

Lexical Overlap

- BLEU: N-gram precision with brevity penalty • METEOR: Synonym-aware with word order consideration
- ROUGE: Recall-oriented content coverage

Translation-Specific

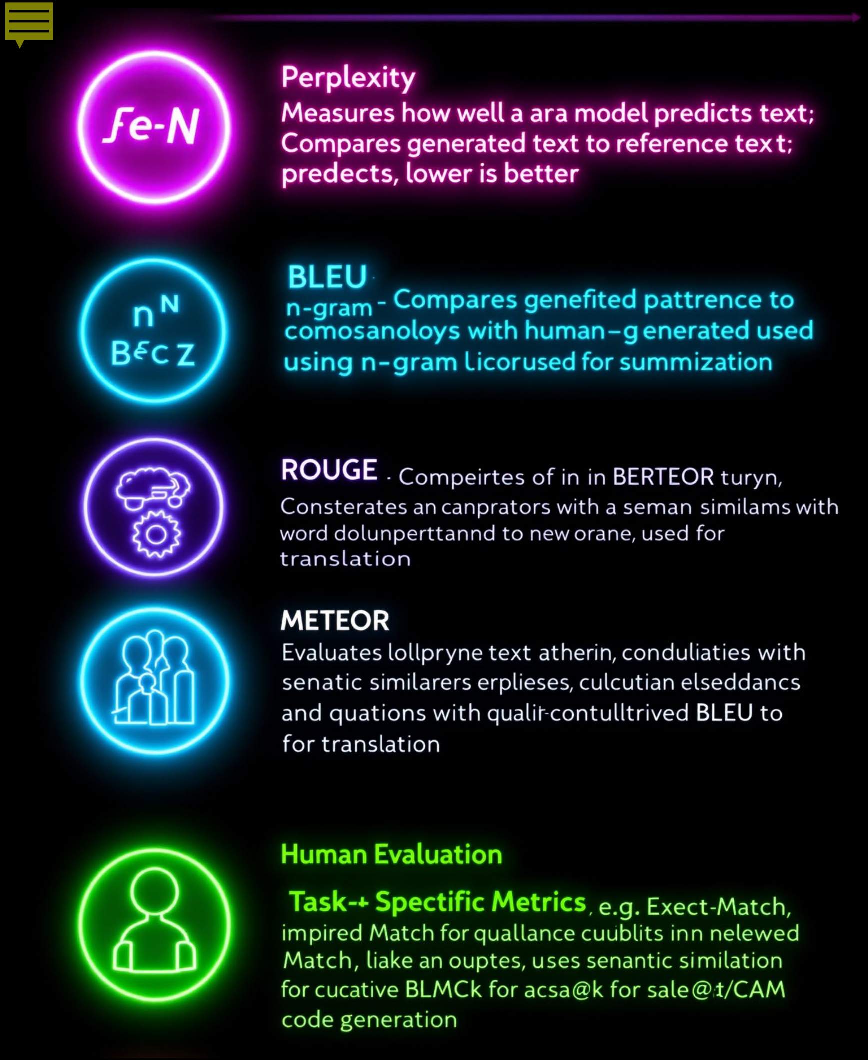
- Adequacy: Meaning preservation accuracy
- Fluency: Target language naturalness
- Multi-reference: Multiple correct translation handling

Primary Use: Source-target preservation, linguistic transformation, meaning fidelity

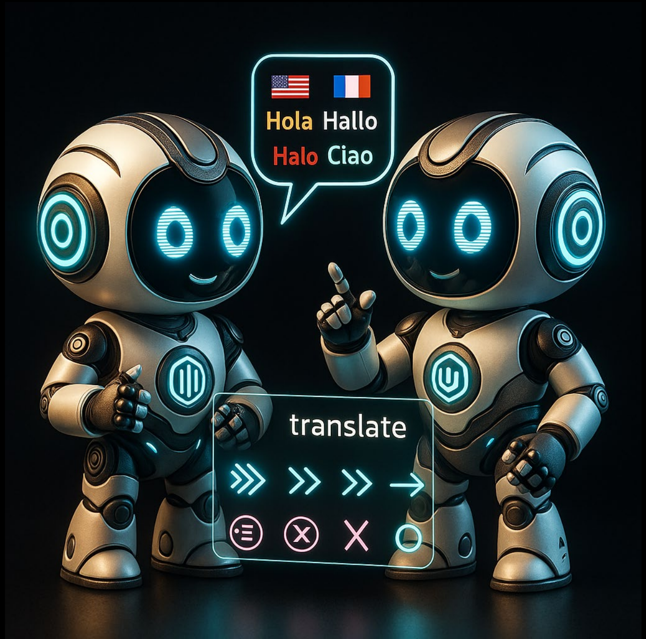
Three Paradigms

Aspect	Deep Learning	LLM Metrics	MT Metrics
Purpose	Training optimization	Generation quality	Meaning preservation
Reference	Ground truth labels	Variable/None	Source text
Creativity	Not applicable	Valued	Discouraged
Subjectivity	Objective	Highly subjective	Moderately objective
Speed	Very fast	Moderate	Fast

[4] [5]



Most used Metrics for MT



BLEU & METEOR

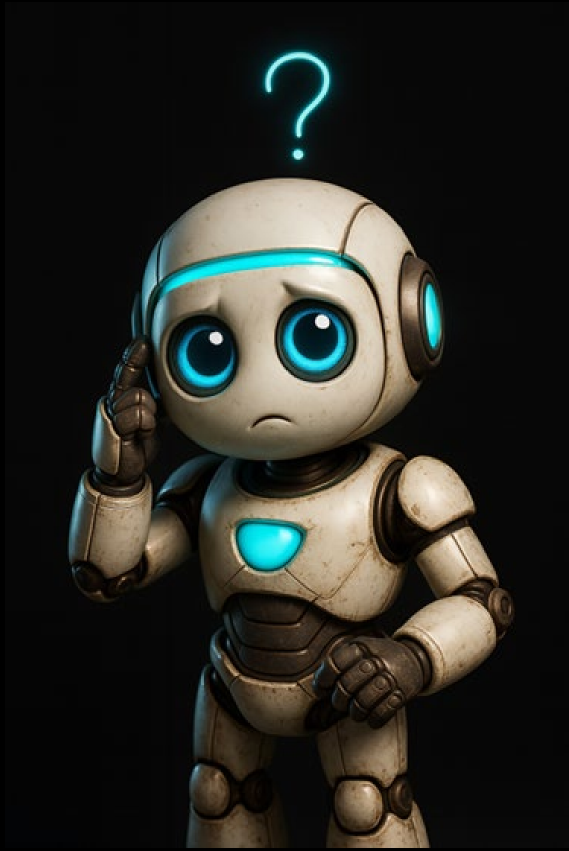
- **BLEU:**
  - n-gram precision metric; compares candidate to reference
  - Scores range from 0 to 1; higher is better
- **METEOR:**
  - aligns unigrams; considers synonyms & stemmed forms
  - Balances precision and recall; penalizes word orde errors

[8]

TER, ROUGE & ChrF

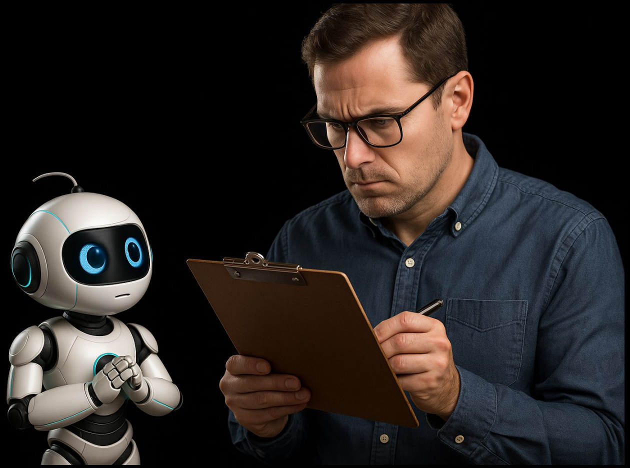
- **TER** (Translation Edit Rate):
  - Counts insertions, deletions & substitutions
  - Measures post-editing effort required
- **ROUGE:**
  - Recall-oriented; measures overlapping n-grams
- **ChrF:**
  - F-score at the character level; captures morphological differences

[9]



Human Evaluation

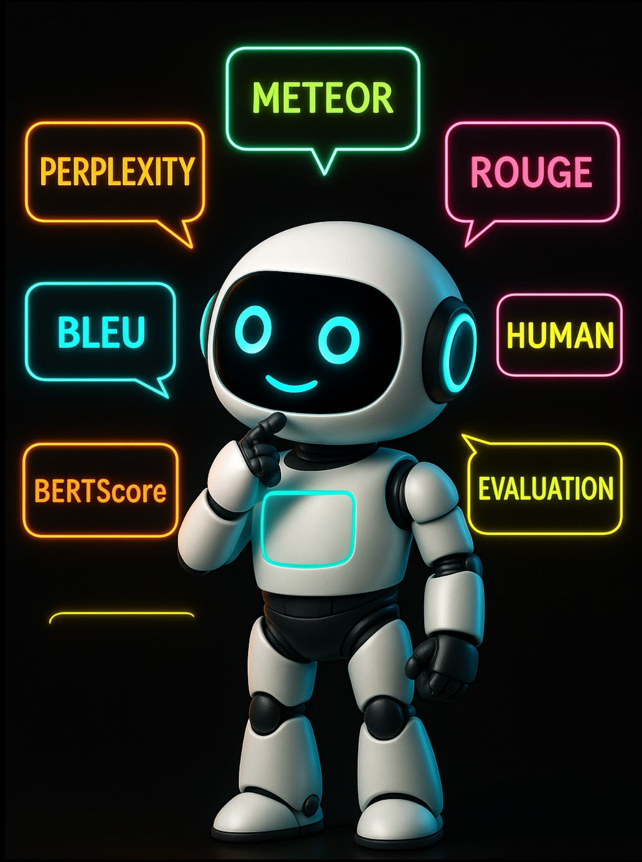
- Fluency: naturalness and readability
- Adequacy: content conveys the same meaning
- Cultural & pragmatic correctness



[10]

Choosing your Metrics

- **BLEU/METEOR:** Quick automatic scoring
- **TER:** Understanding editing effort needed
- **ChrF:** Handling different languages and morphology
- **Human eval:** Final quality check



Speech-to-Speech Translation

ASR
Speech → text

NMT
Translate text

TTS
Text → speech

[11]

Multilingual Models

One model for many languages

- **Multi-language system:** One model handles multiple language pairs instead of separate systems for each pair
- **Zero-shot translation:** Translates between language combinations never seen in training
- **Shared learning benefits:** Learning from many languages improves translation quality across all pairs
- **Resource efficiency**: One model requires less computational power than multiple separate systems

Ethics & Responsible Use

- **Professional Impact:** Effects on human translators
- **Cultural Preservation:** Maintaining linguistic diversity
- **Privacy Concerns:** Sensitive information in translation systems
- **Quality Responsibility**: Who is accountable for translation errors?

Challenges and Limitation

- **Cultural Context:** Idioms, humor, and cultural references
- **Domain Specificity:** Technical, legal, and medical terminology
- **Low-Resource Languages:** Limited training data
- **Bias and Fairness:** Gender, cultural, and social biases

Future Trends

- Handling informal speech, slang & dialects
- Domain adaptation & low-resource languages
- Energy efficiency & computational cost
- Zero-shot and multimodal translation

Key Takeaways

MT evolved from rule-based to statistical to neural

Neural architectures capture context via encoders, decoders & attention

Quality assessed via BLEU, METEOR, TER, ROUGE & ChrF plus human review

Speech translation integrates ASR, NMT & TTS

Applications & future: travel, research, low-resource support & beyond

[12] [13] [14]