A Robust Semantic Text Communication System

Xiang Peng, Zhijin Qin, *Senior Member, IEEE,* Xiaoming Tao, *Senior Member, IEEE,* Jianhua Lu, *Fellow, IEEE,* and Lajos Hanzo, *Life Fellow, IEEE*

***Abstract*—Semantic communication is increasingly viewed as a promising solution to improve the transmission efficiency. However, semantic communications are susceptible not only to physical channel impairments, but also to semantic impairments, which degrade semantic understanding at the receiver and disrupt the associated downstream tasks. Hence, we focus our attention on the robustness of semantic communications against semantic impairments. Specifically, we first categorize textual semantic impairments into three categories based on their sources. Then, we propose a robust deep learning enabled semantic communication system (R-DeepSC) by introducing a semantic corrector for robust semantic encoding so as to facilitate semantic transmission. Moreover, we develop a non-autoregressive version of R-DeepSC, namely NA-RDeepSC, which offers improved inference speed by relying on a non-autoregressive architecture and an adaptive generator embedded into the semantic decoder. NA-RDeepSC performs semantic decoding in parallel, hence reducing the decoding complexity from $O(n)$ to $O(1)$ with a comparable performance to that of R-DeepSC. Our experimental results demonstrate the superior robustness of the proposed R-DeepSC and NA-RDeepSC architectures in eliminating semantic impairments, hence highlighting the significance of this work in advancing the development of robust semantic communications.**

***Index Terms*—Semantic communication, semantic impairments, calibrated self-attention, non-autoregressive, text transmission.**

I. Introduction

In contrast to conventional communications, semantic communications are designed and optimized in ‘semantic space’ for narrowing the semantic discrepancy between the transmitter and the receiver, rather than minimizing the classic symbol error rate [2]. Explicitly, the transmitted content typically represents task-oriented features conveying semantics. Consequently, the optimization objective of semantic communications is no longer the classic bit error rate or symbol error rate, but the fidelity of the semantic information at the receiver. This optimization objective implies that semantic communications are most suitable for scenarios involving either human-to-machine or machine-to-machine communications.

At the time of writing semantic communication systems, [3]–[15] tend to harness the considerable power of deep neural networks (DNNs) for retrieving the transmitted semantic content. DeepSC [3] is a pioneering example of deep learning aided semantic communications, presenting an efficient joint semantic-channel coding architecture conceived for semantic transmission. Lu *et al.* [4] introduced a confidence-based distillation mechanism and a reinforcement learning-powered semantic communication paradigm. Wang *et al.* [5] proposed a knowledge graphs-based semantic communication system, aiming for reducing data transmission volume and improving semantic similarity by optimizing the associated resource allocation strategies.

Apart from text transmission, similar joint semantic communication designs have also been proposed for diverse other applications. For instance, Weng *et al.* [6] developed a semantic speech communication system, while Xie *et al.* [7] designed a bespoke task-oriented multi-user system. Moreover, researchers have also developed various architectures for image and video transmission. More particularly, Huang *et al.* [8] designed a Generative Adversarial Network-based encoder for image coding relying on adaptive bandwidth allocation. Zhang *et al.* [9] exploited a deep reinforcement learning-based resource allocation scheme to reduce the transmission delay. Zhang *et al.* [10] proposed a semantic communication system for flexible code rate optimization to achieve bandwidth efficiency while maintaining transmission quality. Qin *et al.* [11], [12] presented a computing network enbaled semantic communication system for optimizing the computing resources and a generalized semantic communication framework for leveraging the semantics from source and wireless channels. Jiang *et al.* [13] developed a semantic communication system for video conferencing over hostile time-varing channels. Hanzo *et al.* [14] conceived a model-based parametric semantic coding enhancement technique to improve subjective quality and to harness the limited communication resources by prioritizing semantic regions. Xie *et al.* [15] devised a semantic communication system incorporating a memory module for conducting scenario question answering.

Although the aforementioned contributions have succeeded in expanding the range of tasks that semantic communications can perform, the study of their robustness against transmission impairments is still in its infancy. Specifically, the robustness of semantic communications is affected by a pair of impairments. On the one hand, the transmitted signals are corrupted by the inevitable physical channel impairments, such as pathloss, slow and fast fading, dispersion, as well as the noise. These impairments can be mitigated by channel equalization [16] and channel coding [17], while relying on channel estimation [18], which have been extensively investigated.

On the other hand, semantic communications are also contaminated by semantic impairments causing semantic mismatch between the transmitter and the receiver [19]. Fig. 1 illustrates the concept of semantic impairments, which degrade the integrity of semantic communication systems, namely

Part of this work was presented at the IEEE Global Communications Conference 2022 [1]. (Corresponding author: Xiaoming Tao).

Xiang Peng, Zhijin Qin, Xiaoming Tao and Jianhua Lu are with Department of Electronic Engineering, Tsinghua University, Beijing, China. Zhijin Qin is also with the Beijing National Research Center for Information Science and Technology, Beijing, China. Lajos Hanzo is with School of Electronics and Computer Science, University of Southampton, United Kingdom. Email: px21@mails.tsinghua.edu.cn, {qinzhijin, taoxm, lhh-dee}@tsinghua.edu.cn, lh@ecs.soton.ac.uk.

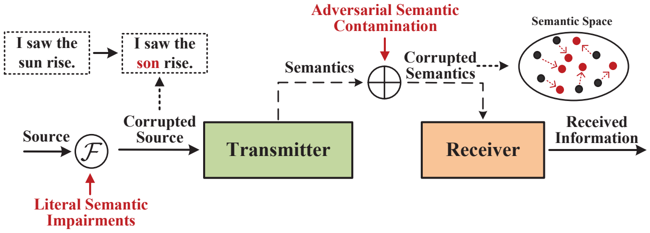


Fig. 1. The semantic impairments in semantic communications.

adversarial semantic contamination and *literal semantic impairments* [20]. *Adversarial semantic contamination* degrades the integrity of semantic communications through semantic channels, while a literal semantic impairment is imposed by corrupted source data. Despite the potentially grave impact of both types of semantic degradations, they have not been extensively studied, hence highlighting the pressing need for further research in this area.

The main focus of this paper is on *literal semantic impairments* of text transmission, which can arise from various sources, including typing errors introduced by users or incorrect recognition by DNN-based systems, such as automatic speech recognition (ASR) algorithms. This type of impairment may result in misspelled or homophonic words that can cause semantic ambiguity. For instance, consider the sentence "I saw the sun rise". If the word 'sun' is misspelled as 'son' due to speech recognition error caused by their similar pronunciation, this imposes semantic impairments by misinterpreting the sentence, potentially leading to erroneous decisions [21].

Conventional communication systems are typically optimized by minimizing the symbol error rate and lack the ability to extract semantics, hence they are vulnerable to such errors. By contrast, semantic communication systems are expected to eliminate semantic impairments and recover the original meaning even from corrupted text, which is made possible by its ability to understand and interpret semantics. However, existing semantic communication systems primarily focus on physical channel impairments, while overlooking semantic impairments and leading to unreliable communications between the transmitter and the receiver [20].

To establish reliable semantic communications, our efforts are dedicated to three aspects: developing robust DNNs that are essential tools for semantic communications, especially in combating adversarial attacks [22]–[24] and literal errors [25]–[30]. Hence they are eminently suitable for designing communication systems that are robust against semantic impairments [1], [20], [31].

1) Robustness Against Adversarial Attacks: The first approach focuses on enhancing the robustness against adversarial attacks, which are perturbations that are intentionally designed to mislead DNNs for producing counter-intuitive predictions. These methods involve designing effective defense strategies. Szegedy *et al.* [22] found that adding invisible perturbations to an image may still deceive a classification model. Goodfellow *et al.* [23] developed a protection mechanism based on a fast gradient method, while Miyato *et al.* [24] proposed a semi-supervised method to defend against adversarial attacks. These methods tend to aim for increasing the resilience of DNNs by adding adversarial examples to the training data. By incorporating these defensive strategies, DNN models become more robust to semantic impairments.

2) Robustness Against Literal Errors: Literal errors may gravely affect the semantic perception of DNNs [32]. To address this issue, designing robust DNNs capable of correcting literal errors is desired. Literal errors typically arise from the following pair of distinct scenarios:

Firstly, errors may have accidentally been made by users during spelling. Numerous studies have focused on eliminating spelling errors. Zhao *et al.* [25] investigated the grammatical error correction capability at the data level by harnessing a dynamic mask for generating 'clean-corrupt' example pairs for training. Zhao *et al.* [26] also introduced a copy mechanism to build a pre-trained model, so as to improve the accuracy of grammatical error correction. Zhang *et al.* [27] proposed a novel detection and correction framework to deal with Chinese literal errors. By applying error correction methods, DNNs can better cope with semantic impairments and perform well in downstream tasks.

Furthermore, literal errors can also be generated due to the limitations of the DNN-based algorithms. For instance, automatic speech recognition algorithms may generate literal errors, due to the limited performance of recognition accuracy, background noise, and the clarity of speech sources [33]. Compared to spelling errors, ASR errors exhibit significant differences in terms of their nature. In addition to misspellings, ASR could also introduce homonym errors, which are caused by the similar pronunciation of words, such as 'sun' and 'son'. To address these challenges, researchers have developed various techniques to reduce the probability of ASR errors. Leng *et al.* [28] proposed an edit alignment method to generate edit labels for 'clean-corrupt' data pairs, which are utilized for training. Zhang *et al.* [29] proposed a dual channel model that leverages both contextual and phonetic information for ASR error correction. Li *et al.* [30] designed the pre-trained BERT-based model and a copy mechanism to eliminate ASR errors. By reducing the impact of semantic impairments in ASR systems, DNNs succeed in reliably interpreting the semantics of speech sources.

3) Robust Design of Semantic Communications: More recently, these successful defense and correction techniques have also been adopted for improving the robustness of communication systems. Peng *et al.* [1] proposed a robust semantic communication system to combat *adversarial semantic contamination* and a specific type of *literal semantic impairments*. Hu *et al.* [20] proposed a robust semantic communication system relying on shared codebooks to tackle both sample-dependent and sample-independent semantic contamination. Sadeghi *et al.* [31] studied the robustness of an end-to-end communication system to physical adversarial attacks and defined a metric termed as the perturbation-to-signal ratio for characterizing the strength of *adversarial semantic contamination*.

Just like any other communication systems, semantic text systems are also prone to semantic impairments [34], which are harder to mitigate than channel impairments. Despite the progress made in addressing the deleterious effects of

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2024.3381950

TABLE I
CONTRASTING OUR CONTRIBUTIONS TO THE LITERATURE

Capability to Withstand	[3]–[15]	[25]–[27]	[28]–[30]	[20], [31]	[1]	Our Work
Physical Channel Impairments	✔			✔	✔	✔
Literal Semantic Impairments by Spelling		✔			✔	✔
Literal Semantic Impairments by ASR			✔			✔
Comprehensive *Literal Semantic Impairments*						✔

semantic impairments, challenges in enhancing the robustness of semantic communication systems persist, including the lack of unified taxonomy and metrics, the absence of semantic impairments datasets for training, as well as the challenges in designing effective yet affordable decontamination modules.

In this context, the existing error correction techniques applied to natural language processing do not consider the realistic constraints of the communication process, while the semantic communication studies have not as yet addressed the grave potential impact of semantic impairments. Therefore, this paper investigates the mechanisms of *literal semantic impairments* and addresses these challenges. Table I boldly contrasts the contributions of this paper against those reviewed above. Our new contributions are further detailed as follows in a point-wise fashion:

- To quantify the semantic impairments that the proposed system can handle, we categorized semantic impairments into three distinct types and developed a new metric termed as *semantic impairment intensity*. Furthermore, we established a semantic impairments dataset having varying *semantic impairment intensity*.
- We developed a robust semantic communication system, referred to as R-DeepSC, which employs a semantic corrector for robust semantic encoding.
- Additionally, we proposed a speedy version, namely NA-RDeepSC, which utilizes an adaptive generator and non-autoregressive architecture for significantly improving the inference speed while maintaining robustness, making it a cost-effective yet efficient solution for online services.

The rest of this paper is organized as follows. Section II introduces the semantic communication system models with particular emphasis on semantic impairments. Section III presents our proposed robust semantic communication system design, while our experimental results are discussed in Section IV. Finally, section V concludes this paper.

II. THE ROBUST SEMANTIC COMMUNICATION MODEL

In this section, we present a robust semantic communication model that is specifically designed for minimizing the effects of semantic impairments in communication channels and propose a novel technique for modeling semantic impairments. Furthermore, we formulate our problem in detail.

A. The Robust Semantic Communication System Model

Fig. 2 portrays the semantic communication system architecture considered, which can handle both physical channel effects and semantic impairments. Denote the input text as $\mathbf{S}$, which is broken into tokens based on the tokenization rules. For instance, when tokenizing the sentence “this is predefined”, the resulting tokenized sequence could be [‘this’, ‘is’, ‘predefined’] or [‘this’, ‘is’, ‘pre’, ‘defined’], depending on the specific tokenization rules used. During the process of tokenization, the collection of all tokens is referred to as the dictionary, denoted as $\boldsymbol{\nu}$. This dictionary serves as the knowledge base in the proposed system, facilitating semantic encoding and decoding. The tokenized sequence can be represented as $\mathbf{S} = \{s_1, s_2, \cdots, s_L\}$, where s_i is the i-th token.

Then, through the One-Hot encoding and the embedding layer, these tokens can be converted into the embedding vector, $\mathbf{E}$, which are represented as

$$\mathbf{E} = f_{\boldsymbol{\gamma}}[f_{\boldsymbol{\nu}}[\mathbf{S}]], \tag{1}$$

where $f_{\boldsymbol{\nu}}[\cdot]$ represents the One-Hot encoder's action associated with the released knowledge base $\boldsymbol{\nu}$ and $f_{\boldsymbol{\gamma}}[\cdot]$ is the embedding layer relying on the trainable parameter set $\boldsymbol{\gamma}$.

Before tranmission, the robust semantic communication system of Fig. 2 must carry out semantic encoding to extract the pertinent semantic features, followed by semantic correction to refine the semantics, and deep learning enabled channel encoding to guard against physical channel impairments, including impairments caused by AWGN and Rician fading channels. Therefore, the transmitted signal, $\mathbf{X}$, is given by

$$\mathbf{X} = f_{\boldsymbol{\varphi}}[f_{\boldsymbol{\lambda}}[f_{\boldsymbol{\eta}}[\mathbf{E}]]], \tag{2}$$

where $f_{\boldsymbol{\varphi}}[\cdot]$ represents the channel encoder's action associated with the trainable parameter set $\boldsymbol{\varphi}$, $f_{\boldsymbol{\lambda}}[\cdot]$ is the semantic corrector having the trainable parameter set $\boldsymbol{\lambda}$, and $f_{\boldsymbol{\eta}}[\cdot]$ is the semantic encoder having the trainable parameter set $\boldsymbol{\eta}$.

The transmitted signal, $\mathbf{X}$, may become distorted by the fading channels and receiver noise. Hence the received signal, $\mathbf{Y}$, can be represented as

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N}_{\boldsymbol{p}}, \tag{3}$$

where $\mathbf{H}$ characterizes the fading channel and $\mathbf{N}_{\boldsymbol{p}} \sim \mathcal{CN}(0, \sigma_n^2)$.

By utilizing a channel decoder, adaptive generator, and semantic decoder, the received text, $\hat{\mathbf{S}}$, can be represented as

$$\hat{\mathbf{S}} = g_{\boldsymbol{\zeta}}[g_{\boldsymbol{\mu}}[g_{\boldsymbol{\delta}}[\mathbf{Y}]], \boldsymbol{\nu}], \tag{4}$$

where $g_{\boldsymbol{\delta}}[\cdot]$ is the channel decoder having the trainable parameter set $\boldsymbol{\delta}$, $g_{\boldsymbol{\mu}}[\cdot]$ is the adaptive generator associated with having the trainable parameter set $\boldsymbol{\mu}$, and $g_{\boldsymbol{\zeta}}[\cdot]$ is the semantic decoder having the trainable parameter set $\boldsymbol{\zeta}$. Specifically, the trainable parameters of our system, including the channel encoder and channel decoder, are optimized and obtained by joint training in an end-to-end manner.

The proposed semantic communication system is designed to enhance robustness against semantic impairments, which is alleviated by introducing a semantic corrector at the transmitter to rectify semantic errors. At the receiver side, an adaptive generator is used for producing the input sequence of the semantic decoder to speed up the decoding process. These modules allow the system to cope with diverse types and degrees of semantic impairments, thereby improving the reliability and efficiency of semantic communications.

B. Semantic Impairments

The semantic impairments, $\mathbf{N}_{\boldsymbol{s}}$, which is considered to be *literal semantic impairments* in the source text, $\mathbf{S}$, may pose challenges for both humans and DNN models. Fig. 3 illustrates two ways of generating semantic impairments, namely by spelling errors during typing and recognition errors generated by deficient DNN models, such as ASR and optical character recognition. Semantic impairments may cause semantic ambiguity and mislead the DNN models. For instance, the misspelling of the word ‘excited’ in the sentence “we are exhausted with this movie” may confuse a sentiment analyzer.

Semantic impairments may be introduced by three operations: replacement, R, deletion, D, and insertion, I. The i-th word of a sentence corrupted by semantic impairments, $\mathbf{N}_{\boldsymbol{s}} = \{N_s^1, N_s^2, \dots, N_s^n\}$, is defined as $\mathcal{F}(N_s^i, e, i)$, which is given by

$$\mathcal{F}(N_s^i, e, i) = \begin{cases} s_i = \{N_s^i\}, & e = R, \\ s_i = \emptyset, & e = D, \\ s_i = \{s_i, N_s^i\}, & e = I, \end{cases} \tag{5}$$

where $\mathcal{F}(\cdot)$ is a semantic impairment simulation function that fits the error distribution of users or incomplete DNN-assisted systems, N_s^i is the corresponding corrupted word of u_i, and e is the error type. For instance, the corrupted text “I saw the son rise” is obtained after applying $\mathcal{F}(son, R, 4)$ function to the uncorrupted sentence “I saw the sun rise”. After applying the function $\mathcal{F}(\cdot)$ to the uncorrupted sentence, $\mathbf{U} = \{u_1, u_2, \dots, u_n\}$, the corrupted text, $\mathbf{S}$, can be obtained.

The semantic impairment simulation function $\mathcal{F}(\cdot)$, illustrates how semantic impairments are generated. The objective of the proposed robust semantic communication system is to mitigate these impairments by approximating the inverse function of the semantic impairment simulation function, denoted as $\mathcal{F}^{-1}(\cdot)$.

C. Problem Formulation

The semantic impairment simulation function has no explicit formula, since the distribution of its input variables may vary in different scenarios, it becomes necessary to formulate the associated problem and devise a solution.

The proposed system takes corrupted text with semantic impairments as its input and generates text without semantic impairments. Moreover, when transmitting over physical channels, the transmitted signal will be subject to the effects of channel noise and fading, as seen in Eq. (3).

The objective of the proposed system is to eliminate semantic impairments in the transmitted text and achieve high-fidelity end-to-end semantic communications, which can be represented as

$$\max_{\mathcal{D}} \mathcal{E}(\mathbf{U}, \hat{\mathbf{S}}), \tag{6}$$

where $\mathcal{E}(\cdot)$ quantifies the semantic similarity between the uncorrupted text and the received text, and $\mathcal{D}$ is the semantic impairment dataset. To address this challenge, we design robust deep learning enabled semantic communication systems to tackle the problem at hand.

III. Proposed Robust Semantic Communication Systems

In this section, we propose a robust deep learning aided semantic communication system, namely R-DeepSC, relying on a semantic corrector for robust semantic encoding. Moreover, we develop a non-autoregressive speedy form of R-DeepSC, termed as NA-RDeepSC, which adopts an adaptive generator to perform semantic decoding at an accelerated inference speed. Additionally, we discuss the model’s implementation in practical scenarios.

A. Robust Semantic Encoding Relying on the Semantic Corrector

Vaswani *et al.* [35] calculate attention scores based on the semantic correlation between tokens, regardless whether they are corrupted or not. By applying these scores to the semantic

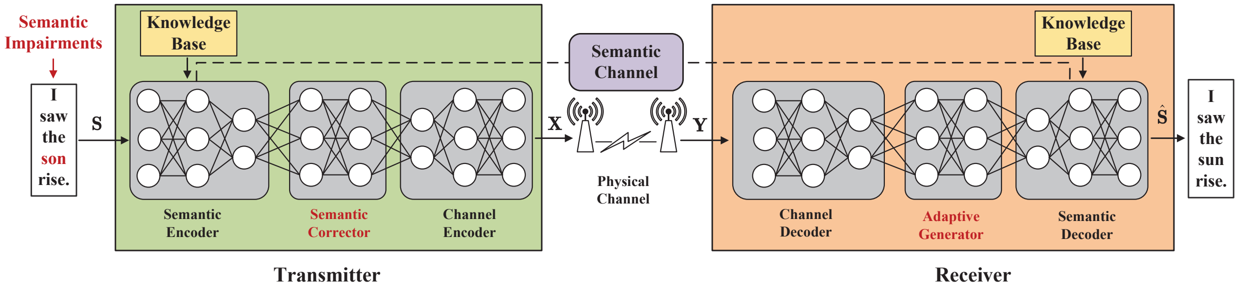


Fig. 2. The robust semantic communication system proposed for mitigating the impact of *literal semantic impairments*. The example in this figure is intended to illustrate the operating principles of our systems, while the proposed systems can handle more complex semantic impairments, encompassing a broader range of forms and higher degrees.

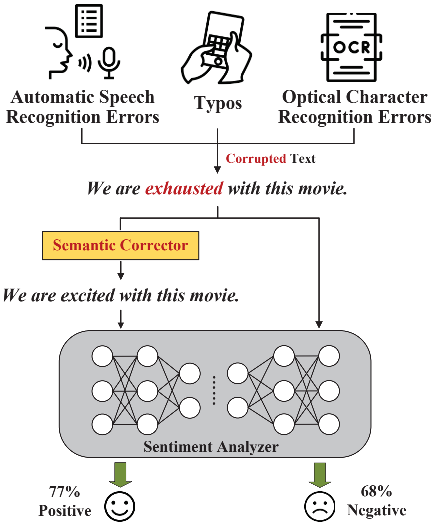

Automatic Speech Recognition Errors
Typos
Optical Character Recognition Errors
OCR
Corrupted Text
We are exhausted with this movie.
Semantic Corrector
We are excited with this movie.
Sentiment Analyzer
77% Positive
68% Negative

Fig. 3. Illustration of the effects of semantic impairments on semantic communications, and the role of the semantic corrector in mitigating these effects.

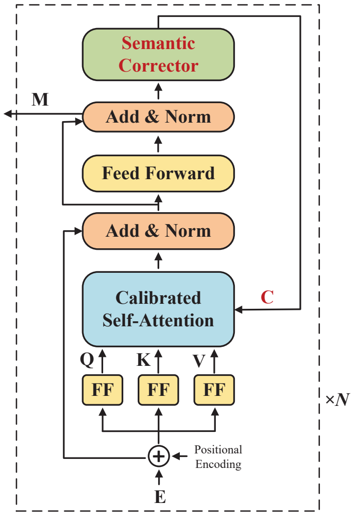


Fig. 4. The robust semantic encoder developed, which relies on a semantic corrector and a calibrated self-attention mechanism.

representations of all tokens, the semantics of the sentence may be obtained. However, if a token is incorrect, its corrupted semantic representation may interfere with the semantics of other tokens, leading to corrupted semantic information. For instance, if an incorrect word is present in the input sequence, such as 'son' instead of 'sun' in "I saw the sun rise", the self-attention mechanism may calculate its representation vector and attention score, which can cause a deviation in the contextual representation of other words and lead to inaccurate model output.

To cope with this problem, we propose a novel semantic encoder which utilizes a semantic corrector and a calibrated self-attention mechanism to eliminate the influence of semantic impairments. For example, in "I saw the son rise", we can adjust the attention score of 'son' to minimize its impact on the contextual representation of other words. This process can help eliminate the interference of corrupted text and improve the accuracy and performance of the Transformer model.

The architecture of the robust semantic encoder developed is shown in Fig. 4. The adopted knowledge base is the dictionary for conducting the Ont-Hot encoding. The extracted semantics, $\mathbf{M}$, is obtained by

$$\mathbf{M} = f_{\boldsymbol{\varrho}}(\mathbf{E}), \tag{7}$$

where $f_{\boldsymbol{\varrho}}(\cdot)$ is the semantic encoder having the trainable parameter set $\boldsymbol{\varrho}$, and $\mathbf{E}$ is the embedding vector obtained with the knowledge base in Equation (1).

A novel semantic corrector is introduced to rectify the corrupted semantics obtained by the semantic encoder, which is the core component of the proposed model, comprising a Gated Recurrent Unit (GRU) [36], a fully connected layer, and a sigmoid activation function. The error probability $\mathbf{P}$ of the tokens, may be represented as

$$\mathbf{P} = f_{\boldsymbol{\epsilon}}(\mathbf{M}), \tag{8}$$

where $f_{\boldsymbol{\epsilon}}(\cdot)$ is the semantic corrector having the trainable parameter set $\boldsymbol{\epsilon}$, and $\mathbf{M}$ is the output of the semantic encoder.

The calibration matrix, $\mathbf{C}$, is formulated as

$$\mathbf{C} = \mathbf{1} - \mathbf{P} \cdot \mathbf{P}^T. \tag{9}$$

The calibration matrix is a weight matrix that adjusts the attention scores of a model to reduce the impact of corrupted text. It assigns smaller weights to the corrupted words, which helps the model better understand the semantic information of the input sequence and enhances its accuracy and performance.

Then, the attention score is calibrated by $\mathbf{C}$ for ensuring that more attention is devoted to uncorrupted tokens. The calibrated attention score, $\mathbf{A_c}$, can be expressed as

$$\mathbf{A_c} = \text{softmax}(\frac{\mathbf{Q} \cdot \mathbf{K}^T}{\sqrt{d_k}} \odot \mathbf{C}), \tag{10}$$

where $\odot$ represents the element-wise product, $\mathbf{Q}$, $\mathbf{K}$, $\mathbf{V}$, d_k are the query, key, value, and the dimension of the encoded semantics.

The calculation process of calibrated self-attention is summarized in Algorithm 1. The value of $\mathbf{C}$ is firstly set as none, which will be updated after passing through the encoder layer of Fig. 2. The semantic error corrector has to be activated $N-1$ times throughout the semantic encoding process, where N is the number of layers.

Furthermore, to train the semantic corrector, a novel loss function, $\mathcal{L}_{SC}(\cdot)$, is developed by relying on the binary cross-entropy loss [37], which is defined as

$$\mathcal{L}_{SC}(\mathbf{P}, \mathbf{L}) = -\sum_i l_i \cdot \log(p_i) + (1 - l_i) \cdot \log(1 - p_i), \tag{11}$$

where $\mathbf{L} = \{l_1, l_2, \ldots, l_n\}$ is the label indicating, whether the token is corrupted or not and l_i can be

$$l_i = \begin{cases} 0, & s_i = u_i, \\ 1, & s_i \neq u_i. \end{cases} \tag{12}$$

Algorithm 1 Algorithm of Calibrated Self-Attention Mechanism

```
Input:
    $\mathbf{Q}$: The query of the input sentence;
    $\mathbf{K}$: The key of the input sentence;
    $\mathbf{V}$: The value of the input sentence;
    $N$ is the number of encoder layers.
Output:
    $\boldsymbol{M}$: The semantic output.
$\mathbf{C} = \mathrm{None}$
for each $i \in [1, \mathrm{N}]$ do
    if $\mathbf{C} = \mathrm{None}$ then
        $\mathbf{M} = \mathrm{softmax}(\frac{\mathbf{Q}\cdot\mathbf{K}^T}{\sqrt{d_k}}) \cdot \mathbf{V}$.
    else
        $\mathbf{M} = \mathbf{A_c} \cdot \mathbf{V}$.
    end if
    $\mathbf{C} = f_{\boldsymbol{\epsilon}}(\mathbf{M}) \cdot f_{\boldsymbol{\epsilon}}(\mathbf{M})^T$.
end for
return $\mathbf{M}$
```

B. Non-Autoregressive Decoder for Inference Acceleration

Although the Transformer of [35] has achieved remarkable performace, the inference time of this autoregressive form has increased substantially due to the complete dependence between tokens. To enhance the inference speed and minimize the communication delay, we concieve a reduced-complexity decoder structure of non-autoregressive form. To realize this goal, the mechanisms of autoregressive and non-autoregressive architecture are analyzed as follows.

An auto-regressive semantic communication system generates the i-th token of the received sentence $\hat{\mathbf{S}}$ as:

$$\hat{s_i} = g_{\boldsymbol{\zeta}}(\mathbf{O}, \hat{s_1}, \hat{s_2}, \cdots, \hat{s}_{i-1}, \boldsymbol{\nu}), \tag{13}$$

where $\mathbf{O}$ is the output of the channel decoder, $g_{\boldsymbol{\zeta}}(\cdot)$ is an autoregressive semantic decoder, and $\boldsymbol{\nu}$ is the knowledge base. Naturally, generating a sequence in an autoregressive manner, which predicts one token at a time based on the previously predicted tokens, has to carry out decoding in series. As a result, the inference time of the autoregressive decoder is on the order of $O(n)$, which incurs an escalating communication delay.

By contrast, a non-autoregressive semantic communication system can transmit data in parallel, because it directly generates a sequence at once. The received text $\hat{\mathbf{S}}$ can be represented as

$$\hat{\mathbf{S}} = g_{\boldsymbol{\pi}}(\mathbf{O}, \mathbf{I}, \boldsymbol{\nu}), \tag{14}$$

where $g_{\boldsymbol{\pi}}(\cdot)$ is the non-autoregressive decoder, which utilizes an independent conditional sequence, $\mathbf{I}$, rather than the tokens, $\hat{s_1}, \hat{s_2}, \cdots, \hat{s}_{i-1}$, generated for conducting semantic decoding. The non-autoregressive architecture is capable of decoding in parallel, hence accomplishing decoding at an inference time order of $O(1)$.

The non-autoregressive architecture is capable of significantly reducing the inference time. However, designing the independent conditional sequence, $\mathbf{I}$, constitutes a critical challenge when establishing a non-autoregressive model. The previously proposed non-autoregressive models rely either on a source-target alignment constraint with fertility [38], or on duration prediction [28] to regulate $\mathbf{I}$. Although these solutions achieve excellent performance, their premise is that the decoder has access to the source text, $\mathbf{S}$, which can then be utilized to build the independent conditional sequence, $\mathbf{I}$, for the semantic decoder.

Unfortunately, this assumption is not applicable to realistic communication scenarios, because as seen in Fig. 2, the semantic decoder receives its input signal, $\mathbf{O}$, from the channel decoder and it can only obtain the source text in case of error-free channel decoding. Therefore, an appropriate conditional sequence must be designed along with the corresponding loss function for training.

In this paper, an adaptive generator, which consists of linear layers, the popular Relu activation function, and the softmax function, is devised for predicting the target length, T, of the input text. The process can be represented as

$$T = g_{\boldsymbol{\mu}}(\mathbf{O}), \tag{15}$$

where $g_{\boldsymbol{\mu}}(\cdot)$ is the adaptive generator module having the trainable parameter set $\boldsymbol{\mu}$.

The k-th token of the input sequence, $\mathbf{I}$, is defined as

$$i_k = \begin{cases} \langle\mathrm{UNK}\rangle, & 0 \leq k \leq T, \\ \langle\mathrm{PAD}\rangle, & k > T, \end{cases} \tag{16}$$

where $\langle\mathrm{UNK}\rangle$ and $\langle\mathrm{PAD}\rangle$ are predefined tokens, indicating that the token is not in the dictionary and the token is used for padding, respectively.

The architecture of the proposed semantic decoder is shown in Fig. 5. Compared to the autoregressive form, the input sequence is no longer constituted by the generated tokens, but by the predicted conditional sequence. Furthermore, since the model predicts in parallel, it is no longer necessary to rely on a mask mechanism, in contrast to the Transformer of [35].

Moreover, the cross-entropy loss function is utilized to develop an adaptive generator loss function, $\mathcal{L}_{AG}(\cdot)$, to regulate the output of the adaptive generator, which is defined as

$$\mathcal{L}_{AG}(T, G) = -G \cdot \log(T), \tag{17}$$

where G is the ground truth for the adaptive generator.

C. Loss Function for Robust Semantic Communications

To allow the whole system to function appropriately, a new loss function is proposed for training the robust semantic communication systems developed, which is given by

$$\begin{aligned} \mathcal{L}_{total} &= \mathcal{L}_{CE}(\mathbf{U}, \hat{\mathbf{S}}) - \alpha \cdot \mathcal{L}_{MI}(\mathbf{X}, \mathbf{Y}) \\ &+ \beta \cdot \mathcal{L}_{SC}(\mathbf{P}, \mathbf{L}) + \gamma \cdot \mathcal{L}_{AG}(T, G), \end{aligned} \tag{18}$$

where $\mathcal{L}_{CE}(\cdot)$ aims for making the uncorrupted text, $\mathbf{U}$, and the received text, $\hat{\mathbf{S}}$, as similar as possible; Furthermore, $\mathcal{L}_{MI}(\cdot)$ maximizes the capacity or the data transmission rate by maximizing the mutual information between the transmitted signal, $\mathbf{X}$, and the received signal, $\mathbf{Y}$; $\mathcal{L}_{SC}(\cdot)$ is the predefined loss used for training the semantic corrector; Finally, $\mathcal{L}_{AG}(\cdot)$

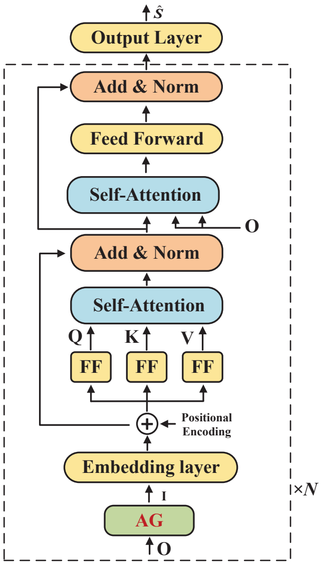


Fig. 5. The semantic decoder developed in NA-RDeepSC.

is the loss employed for training the adaptive generator. The proportions of $\mathcal{L}_{MI}(\cdot)$, $\mathcal{L}_{SC}(\cdot)$, and $\mathcal{L}_{AG}(\cdot)$ in the loss function can be controlled by the positive parameters α, β, and γ.

D. Model Implementation

To enhance the accuracy and efficiency of predictions, we develop a pair of models, namely R-DeepSC and NA-RDeepSC. Specifically, R-DeepSC focuses on transmitting text with a high semantic fidelity by utilizing our robust semantic encoding and autoregressive decoding architecture. Its loss function is composed of the first three terms of $\mathcal{L}_{total}$ in (18), which helps us to optimize the accuracy of the encoding process. By contrast, NA-RDeepSC aims for eliminating the semantic impairments, while maintaining a high inference speed by relying on both our robust semantic encoding and non-autoregressive architecture. Its loss function is $\mathcal{L}_{total}$, which optimizes the overall performance of the model.

The choice between these models depends on the specific task at hand. R-DeepSC is suitable, when the objective is to accurately encode text into a structured representation for transmission. Conversely, NA-RDeepSC is better suited for efficiently decoding structured representations back into text at a superior speed, while maintaining high accuracy. By leveraging the strengths of both R-DeepSC and NA-RDeepSC, our semantic communication solutions are capable of striking a flexible inference accuracy versus speed trade-off.

For time-sensitive scenarios, such as the real-time chat, deploying NA-RDeepSC is advantageous due to its lower inference complexity. Conversely, for the professional document transmission, the R-DeepSC is recommended to ensure high accuracy in error correction. The model selection algorithm is summarized in Algorithm 2.

Algorithm 2 Algorithm of Model Selection

```
Initialization: Load the pretrained R-DeepSC and NA-RDeepSC;
Function: Select model for inference.
Input: Business type $b_t$
if $b_t$ is time-sensitive business then
    Load the paramters of NA-RDeepSC for transmission.
else
    Load the paramters of R-DeepSC for transmission.
end if
```

IV. Numerical Results

In this section, we construct semantic impairments datasets for employment in our experiments. Furthermore, we present our performance metrics, simulation settings, and the experimental results to validate the robustness of the proposed models.

A. Datasets and Baseline Models

We adopt the Europarl corpus dataset [39], which is based on the proceedings of the European Parliament in 11 different languages. The English corpus, which contains 98,751 sentences, is selected as the transmitted data.

Then, a pair of semantic impairments datasets are harvested based on the Europarl dataset. The first dataset is termed as the induced spelling error dataset, which is obtained by randomly sampling the words in the corpus and performing operations based on the predefined transformations of [40] to introduce semantic impairments. The transformation types include substitution, insertion, deletion errors, and verb replacements. We simulate literal errors that may occur in typing by imposing these operations on the corpus, which are determined by sampling a Multinoulli distribution defined in [41]. By adjusting the probability of the above errors and the sampled word index, induced spelling error datasets associated with different levels of *semantic impairment intensity* were collected. The second dataset is referred to as the spontaneous spelling error dataset, which is also based on the Europarl dataset constructed by leveraging the released spelling errror replacement rules [42] relying on the same method.

Moreover, we use the speech-recognition and synthesis based semantic communication system of [43] to transmit elements of the Librispeech [44] dataset over AWGN channels. Briefly, Librispeech is a dataset, which has about 1,000 hours of English speech excerpts, used for conducting ASR tasks. By varying the signal-to-noise ratio (SNR), we obtained an ASR error dataset having different levels of *semantic impairment intensity*.

The proposed models and baseline models are evaluated by relying on these datasets. Details of these datasets are presented in Table II. There are different types of errors in these datasets, which are suitable for comprehensively testing the performance and for yielding reproducible results.

Our proposed system is compared to a range of baseline models. The first one is DeepSC [3], which is a semantic communication system based on deep learning. The second one

TABLE II
DETAILS OF PROPOSED DATASETS

DataSet	Sample Number	Error Types
Spontaneous Spelling Error	93, 809	ReplaceError
Induced Spelling Error	93, 809	InsertError, DeleteError, RepalceError, VerbError
ASR Error	104, 014	WordBoundaryError, SpellingError, GrammaticalError, HomophoneRepalcement

harnesses the SoftMaskedBERT of [27] along with the BERT tokenizer [45] as the semantic codec. The remaining systems use Huffman and low-density parity-check (LDPC) codes with a 0.5 code rate for channel coding, and adaptive modulation [46] techniques for transmission. In AWGN channels, the SNR is stable, but in Rayleigh and Rician fading channels, it can fluctuate significantly. To ensure efficient communication, we use adaptive modulation (AM), which dynamically adjusts the modulation scheme based on the channel conditions to maximize the data rate, while maintaining a low bit error rate. Specifically, we utilize 8-QAM modulation for unfavorable channel conditions, while we employ 16-QAM modulation for good channel conditions. To ensure a fair comparison, DeepSC utilizes the same parameters for training as the proposed R-DeepSC.

B. Simulation Settings

In this experiment, we set the number of layers to 3 and the number of heads to 4. The semantic corrector is set to a gated recurrent unit associated with 128 units and a linear layer, activated by the sigmoid activation. The channel encoder is a dense net having 2 layers, whose hidden dimension is 256 and output dimension of 16. The channel decoder has three layers, with a hidden dimension of 512. The adaptive generator consists of two linear layers and a normalization layer. After passing through the triple-layer semantic decoders, the predicted sequence is generated by the head layer. The details of these settings can be found in Table III.

C. Performance Metrics

Again, in contrast to conventional communication systems, classic metrics, such as the bit-error rate and symbol-error rate, are unable to adequately quantify the performance of semantic communication systems. Instead, we have to consider whether there is a semantic gap between the transmitted and the received text. Hence, we take advantage of the BLEU score [47] and the BERTScore [48] for characterizing the communication performance, while utilizing the *semantic impairment intensity* for quantifying semantic impairments.

1) BLEU Score: The BLEU score utilizes the n-gram matching criterion for evaluating the integrity or intensity of the received text. For example, if we take the sentence "I saw the sun rise", the 1-grams would be 'I', 'saw', 'the', 'sun', and 'rise', while the 2-grams would be 'I saw', 'saw the', 'the sun', and 'sun rise'. We denote the number of the k-th word for the n-gram text by C_k, while the weight of the n-gram precision, and the penalty index by W_n and BP. The BLEU score is formulated as follows

TABLE III
SIMULATION SETTINGS

	Layer Name	Module	Units	Activation
Transmitter	Embeding Layer	Linear	128	None
	Semantic Encoder (× 3)	Transformer Encoder	128 (4 heads)	None
	Semantic Corrector (× 3)	GRU	32 (1 layer)	None
		Linear	1	Sigmoid
	Channel Encoder	Linear	256	ReLU
		Linear	16	None
Receiver	Channel Decoder	Linear	128	ReLU
		Linear	512	ReLU
		Linear	128	None
		LayerNorm	None	None
	Adaptive Generator	Linear	256	ReLU
		LayerNorm	None	None
		Linear	128	Softmax
	Semantic Decoder (× 3)	Transformer Decoder	128 (4 heads)	None
	Head Layer	Linear	Dictionary Size	Softmax

$$\mathrm{BLEU} = \mathrm{BP} \times \exp\Big(\sum_{n=1}^{N} W_n \frac{\sum_i \sum_k \min\left[C_k(R_i)), C_k(T_i)\right]}{\sum_i \sum_k C_k(R_i)}\Big). \tag{19}$$

More particularly, BP is defined as

$$\mathrm{BP} = \begin{cases} 1, & l_R > l_T, \\ e^{1-\frac{l_R}{l_T}}, & l_R < l_T, \end{cases} \tag{20}$$

where l_R corresponds to the length of the received text, and l_T corresponds to the length of the transmitted text. The value of the BLEU score is between 0 and 1, and a higher score implies having a more similar sentence. The BLEU score is efficient, but it only estimates the literal, rather than the semantic difference. As a result, we also harness the BERTScore as the metric of quantifying the semantic similarity between two sentences.

2) BERTScore: The BERTScore quantifies the semantic similarity and applies different weights to words according to their corresponding semantic importance. It was shown in [48] that the semantic similarity assessed by the BERTScore is closely related to human judgements.

We denote the corresponding vector representation of the transmitted text $\mathbf{S}$ by $\langle \mathbf{T_1}, \mathbf{T_2}, \ldots, \mathbf{T_n} \rangle$, and the vector representation of the received text $\hat{\mathbf{S}}$ by $\langle \mathbf{R_1}, \mathbf{R_2}, \ldots, \mathbf{R_m} \rangle$. All these vetors are calculated by the BERT model. The importance weight function $idf(\cdot)$ can be formulated as

$$idf(x) = -\log \frac{1}{M} \sum_{1}^{M} \mathbb{I}(x \in \mathbf{R}^i), \tag{21}$$

where $\mathbf{R}^0, \mathbf{R}^1, \ldots, \mathbf{R}^M$ is the test corpus.

The BERTScore between the transmitted and the received text can be obtained as

$$P_{BERT} = \frac{\sum_{r_i \in \hat{\mathbf{S}}} idf(r_i) \max_{t_i \in \mathbf{S}} \mathbf{T}_i^T \mathbf{R}_i}{\sum_{r_i \in \hat{\mathbf{S}}} idf(r_i)}. \tag{22}$$

Next, the BERTScore is stretched to an expanded range using the following transformation

$$\hat{P}_{BERT} = \frac{P_{BERT} - b}{1 - b}, \tag{23}$$

where b is a scaling factor. The rescaled BERTScore ranges from -1 to 1, and a higher score implies a higher similarity between the pair of input sentences.

3) Semantic Impairment Intensity: Moreover, to quantitatively characterize the semantic impairments, we devise a new metric namely the *semantic impairment intensity* (SII) to quantify the intensity of semantic impairments, which is given by

$$\mathrm{SII} = 1 - \mathrm{BLEU}(\mathbf{S}, \mathbf{U}), \tag{24}$$

where $\mathrm{BLEU}(\cdot)$ is the function quantifying the so-called bilingual evaluation understudy (BLEU) score between the corrupted sentence, $\mathbf{S}$, and the uncorrupted sentence, $\mathbf{U}$. The higher the SII, the stronger the semantic impairments in the source text.

D. System Performance

We conducted comprehensive experiments to validate the performance of the proposed semantic communication systems relying on our semantic impairments datasets.

1) System Performance Versus SNR: Fig. 6 illustrates the performance of our systems for transmission over AWGN channels at various signal-to-noise ratios, in the face of different types of semantic impairments. Specifically, Fig. 6(a), Fig. 6(b), and Fig. 6(c) show the BLEU score of our systems versus the ASR error, spontaneous spelling error, and induced spelling error, respectively. These test datasets have a *semantic impairment intensity* of 0.4, and the models are trained by a combination of three kinds of semantic impairments.

Observe by comparing Figs. 6(a) to 6(c) that our semantic communication systems exhibit lower BLEU scores when tested on ASR error datasets compared to other types of semantic impairments. This result indicates that correcting ASR errors presents the most grave challenge for semantic communication systems. A plausible reason for this is that ASR errors are more complex and have a wider variety of types, making them more difficult to correct. Nonetheless, our solutions still achieve significant improvements in correcting ASR errors, hence they are eminently suitable for practical real-world applications, such as speech recognition.

Furthermore, the results of Fig. 6 suggest that DeepSC struggles to eliminate the semantic impairments inflicted by ASR error datasets, as evidenced by the BLEU scores seen to be lower than 0.6 at high SNRs. This indicates that DeepSC lacks the capability of eliminating semantic impairments without dedicated designs.

By contrast, R-DeepSC efficiently mitigates the semantic impairments imposed by all three datasets, as evidenced by its superior BLEU scores in Fig. 6. This is because R-DeepSC is specifically designed for correcting semantic errors through robust semantic encoding by relying on the semantic corrector and the calibrated self-attention mechanism of Fig. 2.

Similarly, observe in Fig. 6 that the NA-RDeepSC is also capable of mitigating both spontaneous and induced spelling errors. However, it also struggles to correct ASR errors, as evidenced by its lower BLEU scores compared to R-DeepSC. This is because NA-RDeepSC utilizes non-autoregressive decoding, which limits its ability to handle the complex errors inflicted by the ASR datasets.

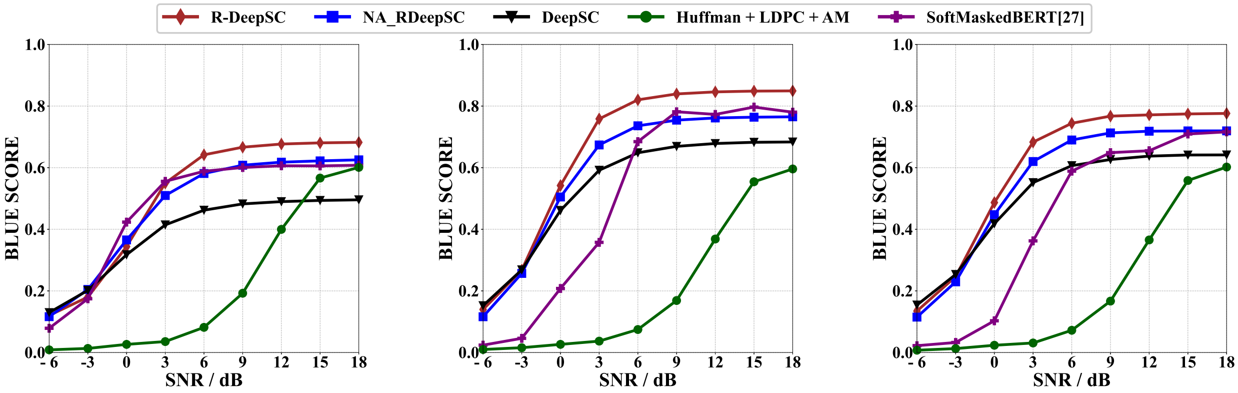
In addition to AWGN channels, we also conducted experiments under Rician fading channels ($k = 1$) associated with various literal errors. The results shown in Fig. 7 exhibit similar trends to those under AWGN channels. Explicitly, the error correction capability quantified in the face of these datasets follows the order of ASR error $<$ induced spelling error $<$ spontaneous spelling error, in line with the gravity of the afflictions experienced.

Furthermore, the performance gap between NA-RDeepSC and R-DeepSC becomes narrower under Rician fading channels compared to AWGN channels. This could be attributed to the fact that Rician fading channels often impose more grave channel impairments at a given SNR. However, the proposed NA-RDeepSC beneficially leverages both the calibrated self-attention mechanism and the non-autoregressive decoding architectures, which allow it to handle the complex errors arising in Rician fading channels more effectively. As a result, NA-RDeepSC achieves more similar transmission performance to that of R-DeepSC under Rician fading channels, despite the complexity of the propagation environment.

2) System Performance Versus SII: To further evaluate the performance of these communication systems, we conducted experiments under various *semantic impairments intensities*, including 0, 0.2, 0.4, 0.6, and 0.8. Fig. 8 shows the performance versus SII at 18 dB for an AWGN channel. The test set is composed of three types of semantic impairments.

The results indicate that both R-DeepSC and NA-RDeepSC outperform the other systems, especially when the SII is greater than 0.2. This demonstrates that R-DeepSC and NA-RDeepSC are capable of supporting robust text transmission.

Fig. 9 shows the performance of semantic communication systems versus the SII under Rician fading channels. The results demonstrate that the semantic fidelity of conventional communication system is significantly degraded in the face of Rician fading channels, while our semantic communication systems achieve superior robustness, as evidenced by both the BLEU score and BERTScore. This is because the proposed NA-RDeepSC and R-DeepSC leverage joint semantic-channel coding methods, allowing them to handle the complex impairments inflicted by Rician fading channels more effectively. Additionally, the models proposed achieve higher semantic

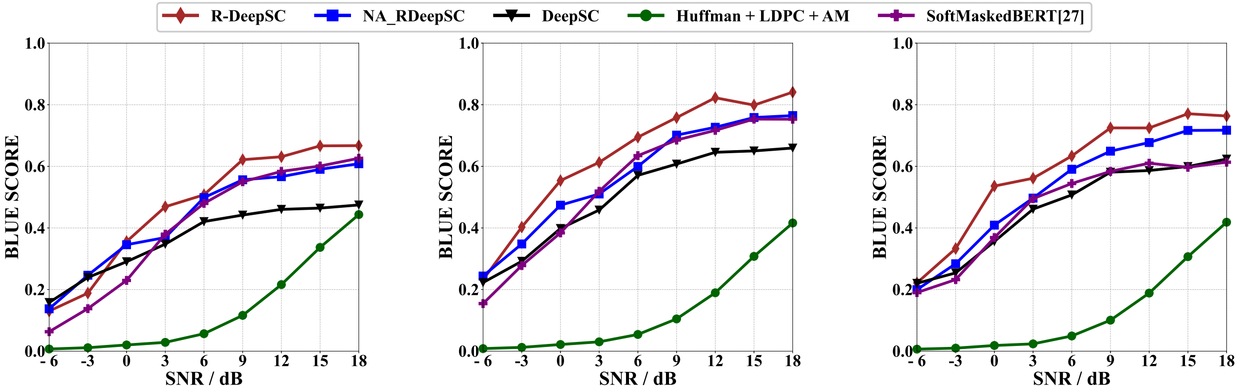


(a) ASR error

(b) Spontaneous spelling error

(c) Induced spelling error

Fig. 6. System performance in AWGN channels versus the SNR with various types of semantic impairments.

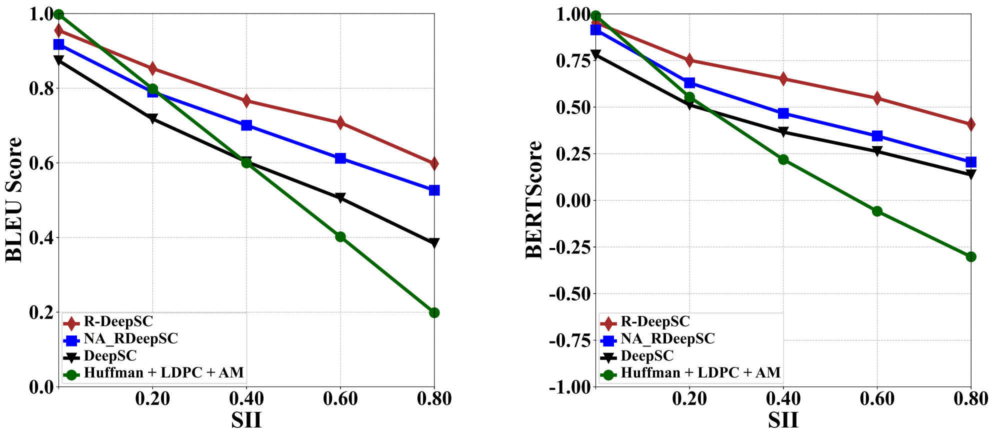


(a) ASR error

(b) Spontaneous spelling error

(c) Induced spelling error

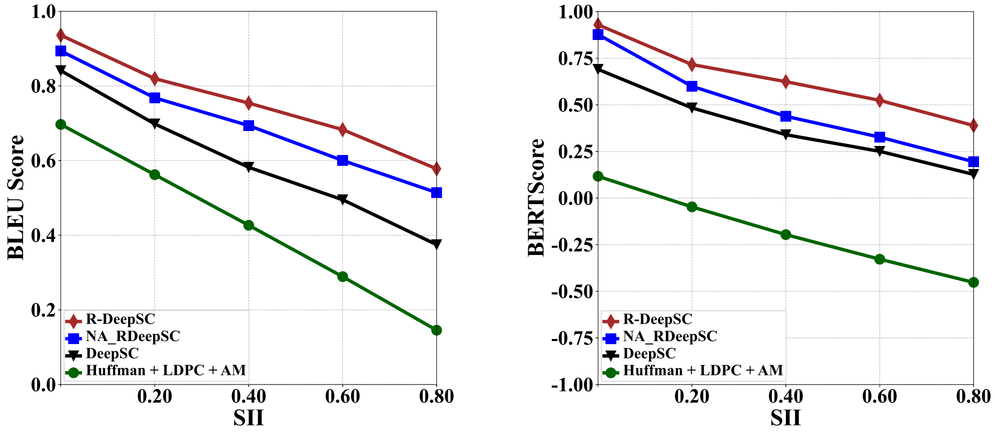
Fig. 7. System performance in Rician fading channels for $k = 1$ versus the SNR with various types of semantic impairments.



(a) BLEU Score

(b) BERTScore

Fig. 8. System performance versus SII under AWGN channels.



(a) BLEU Score

(b) BERTScore

Fig. 9. System performance versus SII under Rician fading channels.

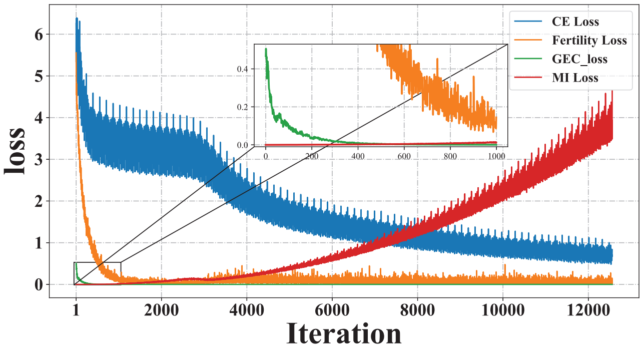


Fig. 10. Loss evolution for NA-RDeepSC for learning rate= 0.0001.

fidelity than DeepSC, even in the face of violently fluctuating SII.

The BLEU score and BERTScore metrics used in this study quantify the text similarity differently, with BLEU evaluating character-level similarity, while the BERTScore measuring semantic similarity. Although they may show a similar tendency in most cases, this is not always the case. For example, in Fig. 8, NA-RDeepSC achieves a similar BLEU score to conventional communication systems using LDPC coding and adaptive modulation associated with SII $= 0.2$, while NA-RDeepSC achieves higher semantic fidelity, as evidenced by its BERTScore. This highlights the importance of considering both metrics for confidently quantifying the performance of semantic communication systems, since they provide different insights in terms of character-level and semantic-level fidelity.

Fig. 10 demonstrates the loss evolution of the proposed NA-RDeepSC. It can be observed that the MI loss keeps on increasing while the other components of the loss function gradually decrease and eventually converge, demonstrating the effectiveness of the system. Table IV shows the transmission results of samples containing different kinds of semantic impairments for SII $= 0.4$, which further demonstrate the effectiveness of our proposed models.

E. *Computational Complexity Analysis*

The proposed NA-RDeepSC exhibits higher inference speed than DeepSC and R-DeepSC as a benefit of its non-autoregressive architecture, which allows for parallel computation and $O(1)$ time complexity. By contrast, DeepSC and R-DeepSC rely on sequential decoding, resulting in a decoding time complexity of $O(n)$. Owing to its novel decoding architecture, NA-RDeepSC maintains a comparable performance, while offering superior inference speed.

Table V offers a comparison of the inference time of the different semantic communication systems. While the inference times of R-DeepSC and DeepSC are comparable due to their autoregressive structures, NA-RDeepSC is significantly faster. Specifically, NA-RDeepSC requires only about 22% of the inference time required by R-DeepSC. This substantial acceleration improves the efficiency of NA-RDeepSC for semantic communications, making it a practical solution for online services.

V. Conclusion

We commenced by categorizing semantic impairments into ASR, spontaneous, and induced spelling errors. To investigate their impact, we have generated semantic impairments datasets and devised the SII metric for our further analysis. We have then conceived the R-DeepSC and NA-RDeepSC schemes. R-DeepSC employs the novel semantic corrector of Fig. 2 to perform robust semantic encoding and an autoregressive scheme for semantic decoding. NA-RDeepSC, which incorporates the R-DeepSC into a non-autoregressive scheme by adopting an adaptive generator to accelerate the inference speed attained. The experimental results demonstrate that both the R-DeepSC and NA-RDeepSC are more robust than the benchmarks, as evidenced by their BLEU score and BERTScore. By applying R-DeepSC and NA-RDeepSC, robust semantic communications can be supported, which could pave the way for their application in real-world scenarios.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC 61925105, 62293484) and Shang-

TABLE IV
TRANSMISSION RESULTS FOR SAMPLES WITH SII $= 0.4$

Text with ASR errors	Though voted same take place at noon todays.
Uncorrupted text	The vote will take place at noon today.
Transmitted by NA-RDeepSC	The vote will take place at noon today.
Transmitted by R-DeepSC	The vote will take place at noon today.
Transmitted by DeepSC	The vote will take place at noon.
Text with spontaneous spelling errors	We shall now priceld to te vite.
Uncorrupted text	We shall now proceed to the vote.
Transmitted by NA-RDeepSC	We shall now proceed to the vote.
Transmitted by R-DeepSC	We shall now proceed to the vote.
Transmitted by DeepSC	We shall now see to the vote.
Text with induced spelling errors	Ut you may uppose I paid no heed.
Correct text	But you may suppose I paid no heed.
Transmitted by NA-RDeepSC	But you may suppose I paid no heed.
Transmitted by R-DeepSC	But you may suppose I paid no heed.
Transmitted by DeepSC	But you may suppose I paid no.

TABLE V
DETAILS OF INFERENCE TIME

Method	Avarge Time (ms / sample)	Total Time (s)
DeepSC	3.321	290.610
R-DeepSC	3.334	291.744
NA-RDeepSC	0.743	65.054

hai Municipal Science and Technology Major Project (Grant No.2018SHZDZX04).

REFERENCES

[1] X. Peng, Z. Qin, D. Huang, X. Tao, J. Lu, G. Liu, and C. Pan, "A robust deep learning enabled semantic communication system for text," in *Proc. IEEE Glob. Commun. (GLOBECOM)*, Dec. 2022, pp. 2704–2709.

[2] Z. Qin, X. Tao, J. Lu, W. Tong, and G. Y. Li, "Semantic communications: Principles and challenges," *arXiv preprint arXiv:2201.01389*, 2021.

[3] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2663–2675, Mar. 2021.

[4] K. Lu, Q. Zhou, R. Li, Z. Zhao, X. Chen, J. Wu, and H. Zhang, "Rethinking modern communication from semantic coding to semantic communication," *IEEE Wireless Commun.*, vol. 30, no. 1, pp. 158–164, 2023.

[5] Y. Wang, M. Chen, W. Saad, T. Luo, S. Cui, and H. V. Poor, "Performance optimization for semantic communications: An attention-based learning approach," in *Proc. IEEE Glob. Commun. (GLOBECOM)*, 2021, pp. 1–6.

[6] Z. Weng and Z. Qin, "Semantic communication systems for speech transmission," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2434–2444, Aug. 2021.

[7] H. Xie, Z. Qin, X. Tao, and K. B. Letaief, "Task-oriented multi-user semantic communications," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 9, pp. 2584–2597, Sep. 2022.

[8] D. Huang, F. Gao, X. Tao, Q. Du, and J. Lu, "Toward semantic communications: Deep learning-based image semantic coding," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 55–71, Jan. 2023.

[9] W. Zhang, H. Zhang, H. Ma, H. Shao, N. Wang, and V. C. M. Leung, "Predictive and adaptive deep coding for wireless image transmission in semantic communication," *IEEE Trans. Wireless Commun.*, vol. 22, no. 8, pp. 5486–5501, 2023.

[10] W. Zhang, Y. Wang, M. Chen, T. Luo, and D. Niyato, "Optimization of image transmission in a cooperative semantic communication networks," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2023.

[11] Z. Qin, J. Ying, D. Yang, H. Wang, and X. Tao, "Computing networks enabled semantic communications," 2023.

[12] Z. Qin, F. Gao, B. Lin, X. Tao, G. Liu, and C. Pan, "A generalized semantic communication system: From sources to channels," *IEEE Wireless Commun.*, vol. 30, no. 3, pp. 18–26, 2023.

[13] P. Jiang, C.-K. Wen, S. Jin, and G. Y. Li, "Wireless semantic communications for video conferencing," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 230–244, Jan. 2023.

[14] L. Hanzo, P. Cherriman, and J. Streit, *Wireless video communications: second to third generation and beyond*. John Wiley & Sons, 2001.

[15] H. Xie, Z. Qin, and G. Y. Li, "Semantic communication with memory," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2658–2669, 2023.

[16] H. Qu, G. Liu, M. A. Imran, S. Wen, and L. Zhang, "Efficient channel equalization and symbol detection for MIMO OTFS systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6672–6686, 2022.

[17] R. G. Maunder, J. Wang, S. X. Ng, L.-L. Yang, and L. Hanzo, "On the performance and complexity of irregular variable length codes for near-capacity joint source and channel coding," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1338–1347, 2008.

[18] X. Guan, Q. Wu, and R. Zhang, "Anchor-assisted channel estimation for intelligent reflecting surface aided multiuser communication," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 3764–3778, 2022.

[19] J. Bao, P. Basu, M. Dean, C. Partridge, A. Swami, W. Leland, and J. A. Hendler, "Towards a theory of semantic communication," in *Proc. IEEE Network Science Workshop*, West Point, NY, USA, Jun. 2011, pp. 110–117.

[20] Q. Hu, G. Zhang, Z. Qin, Y. Cai, G. Yu, and G. Y. Li, "Robust semantic communications with masked VQ-VAE enabled codebook," *IEEE Trans. Wireless Commun.*, pp. 1–10, Apr. 2023.

[21] Y. Zang, F. Qi, C. Yang, Z. Liu, M. Zhang, Q. Liu, and M. Sun, "Word-level textual adversarial attacking as combinatorial optimization," in *Proc. Annual Meeting Assoc. Comput. Linguistics (ACL)*, Online, Jul. 2020, pp. 6066–6080.

[22] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. J. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Banff, Canada, Jan. 2014, pp. 1–10.

[23] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, Mar. 2015, pp. 1–11.

[24] T. Miyato, A. M. Dai, and I. Goodfellow, "Adversarial training methods for semi-supervised text classification," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Toulon, France, Dec. 2017, pp. 1–11.

[25] Z. Zhao and H. Wang, "MaskGEC: Improving neural grammatical error correction via dynamic masking," in *Proc. Assoc. Advancement Artif. Intell. (AAAI)*, vol. 34, Hilton New York Midtown, New York, New York, USA, 04 2020, pp. 1226–1233.

[26] W. Zhao, L. Wang, K. Shen, R. Jia, and J. Liu, "Improving grammatical error correction via pre-training a copy-augmented architecture with unlabeled data," in *Proc. N. Am. Chapter Assoc. Comput. Linguistics (NAACL)*, Minneapolis, Minnesota, USA, Jun. 2019, pp. 156–165.

[27] S. Zhang, H. Huang, J. Liu, and H. Li, "Spelling error correction with soft-masked BERT," in *Proc. Annual Meeting Assoc. Comput. Linguistics (ACL)*, Online, Jul. 2020, pp. 882–890.

[28] Y. Leng, X. Tan, L. Zhu, J. Xu, R. Luo, L. Liu, T. Qin, X. Li, E. Lin, and T.-Y. Liu, "Fastcorrect: Fast error correction with edit alignment for automatic speech recognition," in *Proc. Neural Inform. Process. Systems (NeurIPS)*, vol. 34, Online, May. 2021, pp. 21 708–21 719.

[29] F. Zhang, M. Tu, S. Liu, and J. Yan, "ASR error correction with dual-channel self-supervised learning," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Singapore, May. 2022, pp. 7282–7286.

[30] W. Li, H. Di, L. Wang, K. Ouchi, and J. Lu, "Boost transformer with BERT and copying mechanism for ASR error correction," in *Proc. Int. Joint Conf. on Neural Networks (IJCNN)*, online, 2021, pp. 1–6.

[31] M. Sadeghi and E. G. Larsson, "Physical adversarial attacks against end-to-end autoencoder communication systems," *IEEE Commun. Lett.*, vol. 23, no. 5, pp. 847–850, May. 2019.

[32] M. Shafique, M. Naseer, T. Theocharides, C. Kyrkou, O. Mutlu, L. Orosa, and J. Choi, "Robust machine learning systems: Challenges, current trends, perspectives, and the road ahead," *IEEE Des. Test. Comput.*, vol. 37, no. 2, pp. 30–57, Apr. 2020.

[33] S. Gandhi, P. Von Platen, and A. M. Rush, "ESB: A benchmark for multi-domain end-to-end speech recognition," *arXiv preprint arXiv:2210.13352*, 2022.

[34] K. Lu, R. Li, X. Chen, Z. Zhao, and H. Zhang, "Reinforcement learning-powered semantic communication via semantic similarity," *arXiv preprint arXiv:2108.12121*, 2021.

[35] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Neural Inform. Process. Systems (NeurIPS)*, vol. 30, Long Beach, CA, USA, Dec. 2017, pp. 1–11.

[36] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. Empir. Methods Nat. Lang. Process. (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724–1734.

[37] U. Ruby and V. Yendapalli, "Binary cross entropy with deep learning technique for image classification," *Int. J. Adv. Trends Comput. Sci. Eng*, vol. 9, no. 10, 2020.

[38] J. Gu, J. Bradbury, C. Xiong, V. O. Li, and R. Socher, "Non-autoregressive neural machine translation," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Vancouver, BC, Canada, Feb. 2018.

[39] P. Koehn, "Europarl: A parallel corpus for statistical machine translation," *Mt Summit*, vol. 5, Sep. 2008.

[40] A. Awasthi, S. Sarawagi, R. Goyal, S. Ghosh, and V. Piratla, "Parallel iterative edit models for local sequence transduction," in *Proc. Empir. Methods Nat. Lang. Process. Int. Joint Conf. Nat. Lang. Process. (EMNLP-IJCNLP)*, Hong Kong, China, Nov. 2019, pp. 4259–4269.

[41] C. Robert, "Machine learning, a probabilistic perspective," 2014.

[42] A. Piktus, N. B. Edizel, P. Bojanowski, E. Grave, R. Ferreira, and F. Silvestri, "Misspelling oblivious word embeddings," in *Proc. N. Am. Chapter Assoc. Comput. Linguistics (NAACL)*, Minneapolis, Minnesota, USA, Jun. 2019, pp. 3226–3234.

[43] Z. Weng, Z. Qin, X. Tao, C. Pan, G. Liu, and G. Y. Li, "Deep learning enabled semantic communications with speech recognition and synthesis," *IEEE Trans. Wireless Commun.*, pp. 1–18, Feb. 2023.

[44] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, South Brisbane, Queensland, Australia, Apr. 2015, pp. 5206–5210.

[45] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. N. Am. Chapter Assoc. Comput. Linguistics (NAACL)*, Minneapolis, Minnesota, USA, Jun. 2019, pp. 4171–4186.

[46] K. Zheng and X. Ma, "Designing learning-based adversarial attacks to MIMO-OFDM systems with adaptive modulation," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2023.

[47] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: a method for automatic evaluation of machine translation," in *Proc. Annual Meeting Assoc. Comput. Linguistics (ACL)*, Philadelphia, Pennsylvania, USA, Jul. 2002, pp. 311–318.

[48] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "BERTScore: Evaluating text generation with BERT," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Addis Ababa, Ethiopia, Apr. 2020, pp. 1–43.

Xiang Peng received the B.S. degree from Beijing JiaoTong University and M.S. degree from Tsinghua University. He is pursuing the Ph.D. with the Tsinghua University, Beijing, China. His reseaech interests mainly include signal processing and semantic communicaiton.

Zhijin Qin (Senior Member, IEEE) is an Associate Professor with Tsinghua University, Beijing, China. She was with Imperial College London, London, U.K., Lancaster University, Lancaster, U.K., and Queen Mary University of London, London, from 2016 to 2022. Her research interests include semantic communications and sparse signal processing. She was the recipient of the 2017 IEEE GLOBECOM Best Paper Award, 2018 IEEE Signal Processing Society Young Author Best Paper Award, 2021 IEEE Communications Society Signal Processing for Communications Committee Early Achievement Award, 2022 IEEE Communications Society Fred W. Ellersick Prize, and 2023 IEEE ICC Best Paper Award. She was the Guest Editor of IEEE Journal on Selected Areas in Communications (JSAC) Special Issue on Semantic Communications and Area Editor of IEEE JSAC Series. She was also the Symposium Co-Chair of IEEE GLOBECOM 2020 and 2021. She is an Associate Editor for IEEE Transactions on Communications, IEEE Transactions on Cognitive Networking, and IEEE Communications Letters.

Xiaoming Tao (Senior Member, IEEE) received the Ph.D. degree in information and communication systems from Tsinghua University, Beijing, China, in 2008. She is currently a Professor with the Department of Electronic Engineering, Tsinghua University. Prof. Tao was a workshop General Co-Chair for IEEE INFOCOM 2015 and the volunteer leadership for IEEE ICIP 2017. Since 2016, she has been the Editor of the Journal of Communications and Information Networks and China Communication. She was the recipient of the National Science Foundation for Outstanding Youth (2017C2019) and many national awards, which include the 2017 China Young Women Scientists Award, 2017 Top Ten Outstanding Scientists and Technologists from China Institute of Electronics, 2017 First Prize of Wu Wen Jun AI Science and Technology Award, 2016 National Award for Technological Invention Progress, and 2015 Science and Technology Award of China Institute of Communications.

Jianhua Lu (Fellow, IEEE) received the B.S. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1986 and 1989, respectively, and the Ph.D. degree in electrical and electronic engineering from the Hong Kong University of Science and Technology, Hong Kong, in 1998. Since 1989, he has been a Professor with the Department of Electronic Engineering, Tsinghua University. He is currently a Vice President of the National Natural Science Foundation of China. He has authored or coauthored more than 300 referred technical papers published in international renowned journals and conferences and more than 80 Chinese invention patents. His research interests include broadband wireless communications, multimedia signal processing, and satellite communications. He was the recipient of the Best Paper Awards at the IEEE ICCCS 2002, China Comm. 2006, IEEE Embedded-Com 2012, IEEE WCSP 2015, IEEE IWCMC 2017, and IEEE ICNC 2019. Prof. Lu was the Editor of IEEE Transactions on Wireless Communications from 2008 to 2011, and Program Committee Co-Chair, and a TPC Member of many international conferences. He is also the Editor-in-Chief of China Communications. He is a Member of Chinese Academy of Sciences.