When Meaning Meets Purpose: A Survey on Semantic and Task-Oriented Communications

Quang Tuan Do[a], Tung Son Do[a], Thanh Phung Truong[a], Dongwook Won[a], Nhu-Ngoc Dao[c], Wonjong Noh[b,*], Sungrae Cho[a,*]

[a] *Department of Computer Science and Engineering, Chung-Ang University, Seoul, 06974, South Korea*
[b] *School of Software, Hallym University, Chuncheon, 24252, South Korea*
[c] *Department of Computer Science and Engineering, Sejong University, Seoul, 05006, South Korea*

Abstract

As wireless systems evolve toward sixth-generation (6G) networks, traditional bit-centric approaches cannot support intelligent applications that require meaning-aware and effectiveness-aware transmission. This survey examines semantic communication (SC) and its specialized subset, task-oriented semantic communication (ToSC), which prioritizes task-specific utility over general semantic completeness. Unlike prior surveys that treat this distinction descriptively, we synthesize the information bottleneck framework with empirical evidence from the surveyed literature to characterize the qualitative regimes under which SC and ToSC diverge or converge, and to distill design heuristics. An architectural taxonomy along functional axes (encoding strategy, transmission mode, task-relevance mechanism, and adaptivity) captures the recurring mechanisms shared across implementations. We further analyze the deep learning model families (autoencoders, convolutional and graph neural networks, Transformers, and large language and multimodal models) that implement semantic extraction, compression, and task reasoning. A structured evaluation framework addresses metric definitions and the conflicts and trade-offs between semantic fidelity, task effectiveness, and communication efficiency. Across ten application domains, we identify domain-specific risk factors, unresolved limitations, and four cross-domain weakness patterns: knowledge dependency as the most pervasive bottleneck, weak-baseline comparisons that inflate reported gains, brittleness in safety-critical deployments, and inconsistent evaluation methodology that prevents meaningful cross-domain comparison. Open challenges are consolidated into a dependency-ordered research roadmap that prioritizes foundational theory, standardized benchmarks, and scalable knowledge base management. This survey positions SC as the overarching paradigm for 6G information-directed communications, with ToSC as its action-driven specialization, and offers practical criteria for deciding when each applies and how to evaluate it.

Keywords: Semantic communication, task-oriented communication, deep learning, joint source-channel coding, large language models, sixth-generation wireless networks.

1. Introduction

1.1. Background and Motivation

Sixth-generation (6G) wireless systems are expected to support applications that impose stringent requirements on latency, bandwidth, and energy efficiency, ranging from extended reality and autonomous systems to large-scale Internet of Things (IoT) and smart healthcare [1, 2, 3]. Traditional communication systems, designed for accurate bit-level transfer, are ill-suited to these demands: optimizing for symbol fidelity while disregarding the meaning, context, or task-relevance of information leads to redundant transmissions and inefficient resource usage in intelligent,

*Corresponding author.
Email addresses: dqtuan@uclab.re.kr (Quang Tuan Do), tsdo@uclab.re.kr (Tung Son Do), tptruong@uclab.re.kr (Thanh Phung Truong), dwwon@uclab.re.kr (Dongwook Won), nndao@sejong.ac.kr (Nhu-Ngoc Dao), wonjong.noh@hallym.ac.kr (Wonjong Noh), srcho@cau.ac.kr (Sungrae Cho)

context-aware systems [4, 5, 6]. This qualitative gap, where lowering bit error rates does not improve semantic relevance or downstream effectiveness, motivates rethinking the design objective: communication systems should be aware of semantics and their pragmatic use [7, 8].

1.2. Semantic Communication and Task-Oriented Specialization

Semantic communication (SC) redefines the core objective of communication, moving from the accurate replication of symbols from source to destination toward conveying the sender's intended meaning or semantic content. Although the concept has been explored for decades, recent developments in artificial intelligence (AI) and machine learning offer advanced tools required to extract, represent, and interpret meaning from complex and heterogeneous data, reigniting interest in SC and accelerating its progress. This approach draws inspiration from the efficiency of human communication, which emphasizes intent and context [9, 10]. SC shifts the design question from "how to transmit data" to "what to transmit": by selectively conveying only semantically relevant information, SC reduces redundant communication, conserves bandwidth, and lowers latency [11, 8].

Within this broader paradigm, task-oriented semantic communication (ToSC) is a specialization of SC in which the communication objective is to enable a specific task outcome at the receiver. Success is therefore measured by task-level performance metrics, such as classification accuracy, detection probability, control reward, decision latency, or mission success rate rather than symbol- or bit-level fidelity. ToSC extracts and transmits only the semantic information that is causally useful for the task under resource constraints, while suppressing irrelevant semantics. Early formulations include Xie et al.'s multi-user visual question answering (VQA) system [12] and Kang et al.'s personalized saliency encoder [13]; broader treatments of the ToSC design space appear in [14, 15].

Consider wireless image transmission. Under general reconstruction-oriented SC, the receiver aims to reconstruct a perceptually faithful image for human viewing or broad reuse; background and fine-grained texture must be preserved, and success is measured by perceptual fidelity (e.g., PSNR/SSIM) and quality-of-experience. Under action-oriented ToSC, the receiver's goal may be classification or detection; only class- or target-discriminative features are essential, permitting the encoder to discard details that do not affect task performance. Communication success is measured by task metrics (e.g., top-1/top-5 accuracy, F1-score, false-alarm/miss rates) and end-to-end decision latency under bandwidth/energy constraints [14, 15]. This contrast clarifies that general SC prioritizes reusable meaning, whereas ToSC prioritizes pragmatic effectiveness for a specific action.

This meaning- and action-centric perspective calls for theoretical frameworks and metrics that extend beyond bit error rate to include semantic fidelity for general SC and pragmatic success for ToSC. In this survey, we treat SC as the overarching paradigm guiding information-directed communications for 6G and beyond, and ToSC as a key specialization tailored to action-driven systems.

1.3. Contributions and Organization

This survey provides a systematic examination of SC with specialized emphasis on ToSC, presenting both theoretical foundations and practical implementations. The key contributions of this work are as follows:

1. **Theoretical Foundations Traced to Their Design Consequences:** We trace the theoretical lineage showing that knowledge-base mismatch degrades semantic channel capacity independently of physical channel quality, synthesizing the progression from classical information theory through semantic and strongly semantic information theories, Bao et al.'s semantic communication theory, and goal-oriented communication, and connecting each phase to its practical implications for modern ToSC system design.
2. **A Unified SC/ToSC Framework Stating When the Distinction Matters:** Unlike prior surveys that treat the SC/ToSC distinction as a labeling convention, we synthesize the information bottleneck (IB) framework (originating with Tishby et al. [26], mapped to SC by Beck et al. [27]) with empirical evidence from the surveyed literature to characterize the qualitative regimes under which SC and ToSC diverge or converge. Concrete examples demonstrate the conditions under which SC-optimized systems underperform at tasks and vice versa, and three design principles translate these conditions into practical engineering criteria.
3. **Evaluation Metrics and Their Conflicts:** Metrics are organized by Weaver level (Level B fidelity for SC, Level C effectiveness for ToSC, plus cross-level efficiency metrics). We analyze the conflicts and trade-offs between metric categories, including Level B vs. Level C, effectiveness vs. efficiency, and efficiency vs. fidelity, and provide practitioner guidance for metric selection under competing objectives.

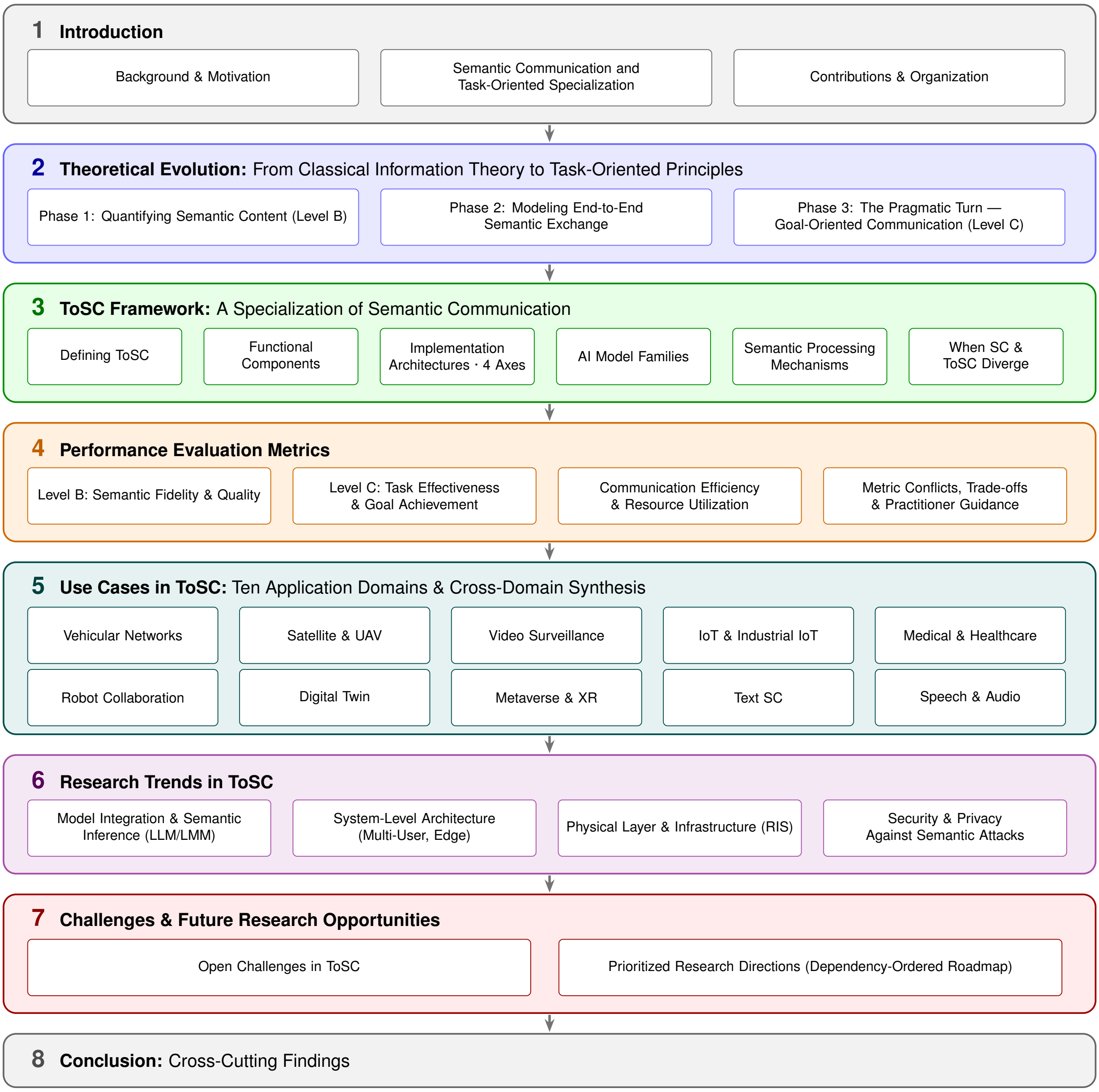

Figure 1: Overall structure of this survey.

4. **Ten Domains Distilled into Four Weakness Patterns:** From ten application areas we distill four recurring weakness patterns not systematically documented in prior surveys (developed in Section 5 and mapped in Fig. 7), each domain carrying a critical assessment of limitations, negative results, and unresolved challenges.
5. **Research Trends and Prioritized Research Directions:** Five research priorities, ranked by the number of downstream barriers each unblocks, distill open challenges from research trends spanning system intelligence (large language model / large multimodal model (LLM/LMM) integration, intent-based and explainable systems), system architecture (multi-user, cooperative agents, edge computing), and physical infrastructure (task-aware design, reconfigurable intelligent surface (RIS)-assisted systems).

These contributions differentiate this survey from prior SC/ToSC surveys along the axes made explicit in Table 1. Rather than cataloging schemes by application domain or by neural-network backbone (the organizations that dominate

Table 1: Positioning of This Survey Relative to Prior Work

Survey	Theory	Architecture	Metrics	Applications	Critical synthesis	Primary focus
[16] (2024)	∘	∘	∘	•	×	Broad SC overview
[17] (2023)	•	•	∘	∘	×	SC enabling-technique taxonomy
[18] (2023)	∘	•	×	∘	×	Context-based design framework
[19] (2023)	∘	×	•	∘	×	SC/GOC performance metrics
[20] (2024)	∘	•	×	∘	∘	SC security threats and countermeasures
[4] (2024)	•	×	∘	×	×	Mathematical foundations
[21] (2026)	•	•	∘	•	∘	Contemporary SC (Theory of Mind, generative AI, JSCC)
[22] (2025)	×	∘	×	∘	∘	Secure SC across the life cycle
[23] (2026)	∘	•	×	∘	×	Semantic radio access networks
[24] (2026)	∘	•	×	•	×	Generative diffusion for wireless
[25] (2026)	∘	•	∘	•	∘	Generative-AI-enabled SC
[14] (2024)	•	∘	∘	∘	∘	Goal-oriented SC: theory, techniques, and challenges
[15] (2023)	•	•	∘	∘	∘	Context/semantics/task-oriented foundations (tutorial)
This work	•	•	•	•	•	SC/ToSC distinction conditions, metric-conflict analysis, cross-domain synthesis, KB stress-test, prioritized roadmap

Coverage (operational scoring, per aspect): • = a dedicated treatment with comparative synthesis across multiple works; ∘ = the aspect is addressed but only descriptively or for isolated schemes; × = absent or mentioned only in passing. "Critical synthesis" denotes the presence of explicit cross-work comparative analysis organized around stated trade-off axes (rather than per-topic description), independent of which specific trade-offs are examined; a • rating additionally requires that this comparative synthesis span the majority of the four content dimensions (theory, architecture, metrics, and applications) rather than a single sub-area. The • cross-work analyses scored for **This work** are located in the theoretical-lineage synthesis (Section 2), the architectural taxonomy along stated functional axes (Section 3.3, Table 3), the metric-conflict analysis (Section 4), the cross-domain weakness synthesis (Section 5, Fig. 7), and the prioritized roadmap (Section 7); all scores reflect the authors' assessment of each comparator's cited version.

earlier surveys), we organize implementations along four *functional axes* (encoding strategy, transmission mode, task-relevance mechanism, and adaptivity) and overlay a complementary AI-model-family lens (Section 3.4), so that recurring design mechanisms become visible across otherwise dissimilar systems. Moreover, we ground the SC/ToSC distinction in the information-bottleneck framework and the surveyed empirical evidence to state *when* it matters.

Among the most closely related recent surveys, the survey of semantic radio access networks [23] offers the most systematic treatment of how semantic communication integrates into the radio access network and protocol stack, but confines its analysis to that layer and so neither derives when the SC/ToSC distinction matters nor analyzes inter-metric conflict. The survey of generative diffusion for wireless [24] catalogs diffusion-model techniques across the wireless protocol stack more comprehensively than any other, yet treats diffusion as a model family rather than as a decoding strategy spanning Weaver Levels B and C (Section 6.1). The dedicated survey of generative-AI-enabled SC by Liang and Li [25] likewise indexes the field by generative architecture, communication modality, and application task, not by the communication objective the generated output serves. The contemporary-SC survey of Nguyen et al. [21] comes closest by aggregate coverage in Table 1 and spans the broadest range of emerging technique families, from theory of mind and generative AI to joint source-channel coding (JSCC); its organizing axis, however, remains the technique itself, leaving open the condition-based synthesis of *when* SC and ToSC diverge and the inter-metric conflict analysis developed here. This combination of a structured cross-survey coverage matrix, a mechanism-level architectural taxonomy, and a critical, condition-based synthesis is, to our knowledge, not jointly offered by existing surveys.

The remainder of this paper is organized as follows (see Fig. 1 for the overall structure). Section 2 presents the theoretical evolution from classical information theory to goal-oriented communication. Section 3 introduces the ToSC framework, including a critical analysis of when the SC/ToSC distinction matters. Section 4 establishes the evaluation framework with metric conflict analysis. Section 5 examines ten application domains with per-domain critical assessments and a cross-domain synthesis. Section 6 surveys research trends. Section 7 discusses challenges and presents prioritized research directions. Section 8 concludes with cross-cutting findings. To contextualize these contributions, Table 1 identifies the specific analytical gaps in prior surveys that this work addresses, focusing on the *types of analysis* that remain missing, not topical coverage breadth. We also distinguish our scope from the complementary view of communication as part of integrated sensing, computation, and communication: whereas recent surveys of edge perception organize the field around how the network edge senses and processes its physical environment [28], this survey centers on the communication objective itself (what meaning or task outcome the transmitted representation must preserve), treating sensing and computation as sources of semantics rather than the organizing axis. Our review covers the semantic- and task-oriented-communication literature available through mid-2026.

Notation. Throughout this paper, X denotes the source data, Z denotes the compressed semantic representation, T denotes the task variable (e.g., a class label), and Y denotes the channel output. $I(\cdot;\cdot)$ denotes mutual information, $H(\cdot)$ denotes Shannon entropy, and $H(\cdot|\cdot)$ denotes conditional entropy, all in the standard information-theoretic sense; $\overline{H}_s(\cdot)$ denotes Bao et al.'s model-theoretic *semantic* entropy (the average logical information of a message under a given knowledge base and inference procedure), distinct from the Shannon entropy $H(\cdot)$. In the IB formulation (Section 3.6.1), $\beta > 0$ controls the compression-relevance trade-off. K_s and K_d denote the source and destination knowledge bases, respectively. $\langle W, K, I, M \rangle$ refers to Bao et al.'s 4-tuple system model (world model, background knowledge, inference engine, message interpreter) introduced in Section 2; the bare I in this tuple is Bao et al.'s inference component, distinct from the mutual-information operator $I(\cdot;\cdot)$ above.

2. Theoretical Evolution: From Classical Information Theory to Task-Oriented Principles

The development of ToSC builds upon a theoretical lineage that progressively extended communication theory beyond symbol transmission toward meaning and purpose. The semiotic foundations of Morris [29] distinguish three dimensions of signs: syntactics (formal structure), semantics (meaning conveyed), and pragmatics (effect on interpreters), which map directly onto Weaver's three communication levels [30], illustrated in Fig. 2:

- **Level A (Technical):** Accuracy of symbol transmission, the domain of Shannon's Classical Information Theory (CIT) [31].
- **Level B (Semantic):** Precision with which transmitted symbols convey intended meaning.
- **Level C (Effectiveness):** Impact of received meaning on achieving desired behavior [32].

CIT provides the mathematical foundation for Level A but explicitly abstracts away Levels B and C. This abstraction is adequate for bit-pipe communication but creates a fundamental mismatch for intelligent 6G applications where the *relevance* and *utility* of transmitted information matter more than its bit-level fidelity. Bridging this gap has been the central theoretical challenge driving SC research, progressing through three distinct phases: quantifying semantic content (Level B theories), modeling semantic exchange in communication systems, and formalizing goal-directed communication (Level C theories). We summarize these phases below and compare them in Supplementary Table S1; readers seeking detailed mathematical treatments are referred to [4, 17]. Fig. 3 traces this progression across the three conceptual phases, mapping each theoretical milestone to its enabling contribution in modern ToSC practice.

2.1. Phase 1: Quantifying Semantic Content (Level B)

The earliest attempts to formalize semantic information (SI) emerged from Bar-Hillel and Carnap's Semantic Information Theory (SIT) [33], which defined SI as the logical content implied by a statement within a formal language model. A statement that narrows the set of possible states carries more SI than a vague one, analogous to Shannon entropy but operating over logical rather than probabilistic spaces. However, SIT suffered from the Bar-Hillel–Carnap paradox: contradictions, being maximally "informative" in a logical sense, received the highest SI values despite being false [33].

Floridi's Strongly Semantic Information Theory (SSIT) [34] resolved this by incorporating truth values directly into SI quantification. SSIT defines a discrepancy function $f(s)$ measuring how far a statement s deviates from the actual state of affairs, and derives informativeness as $g(s) = 1 - f(s)^2$. This formulation ensures that false statements carry negative informational value, aligning SI with the practical intuition that only truthful content supports effective decision-making. Subsequent extensions incorporating "truthlikeness" [35, 36] further refined the granularity of this analysis. Despite these advances, both SIT and SSIT remain limited to logical propositions and do not address how SI propagates through a communication system, a gap that motivated the next theoretical phase. However, the core

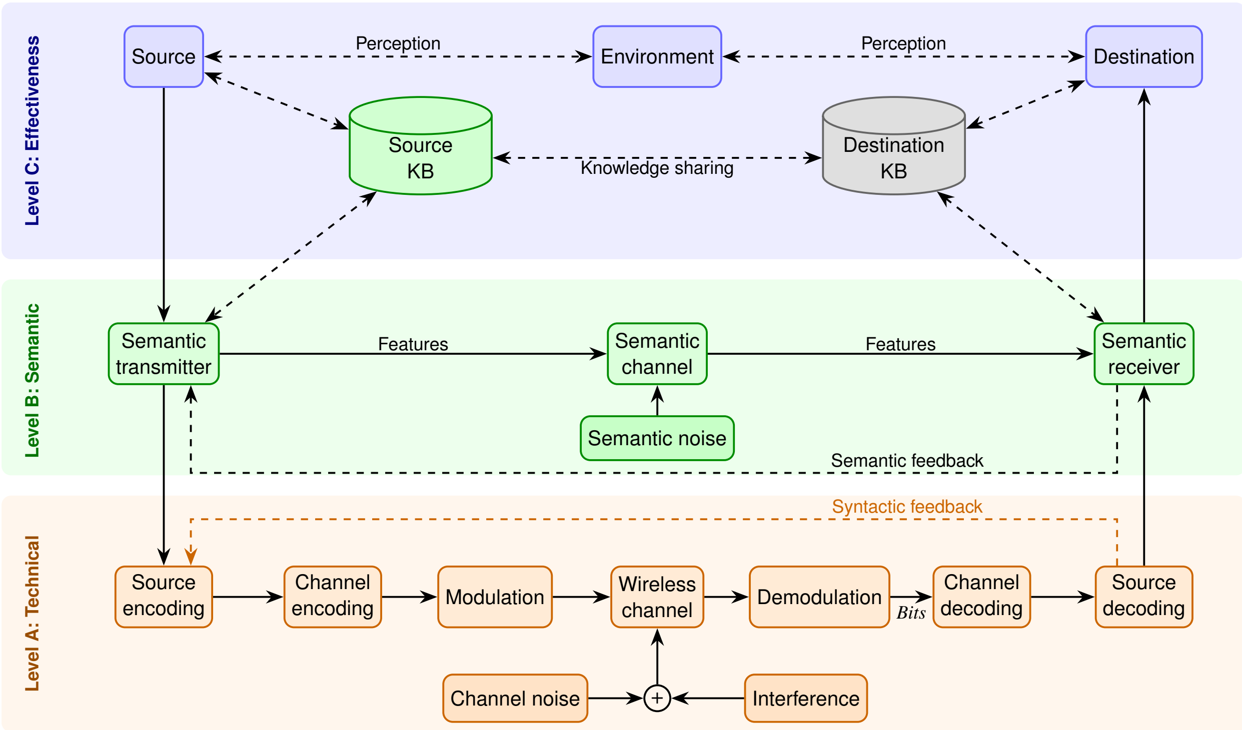

Figure 2: Weaver's three levels of communication (Level A: Technical, Level B: Semantic, and Level C: Effectiveness). The mapping of these levels to the performance metrics is discussed in Section 4; ToSC primarily targets Level C.

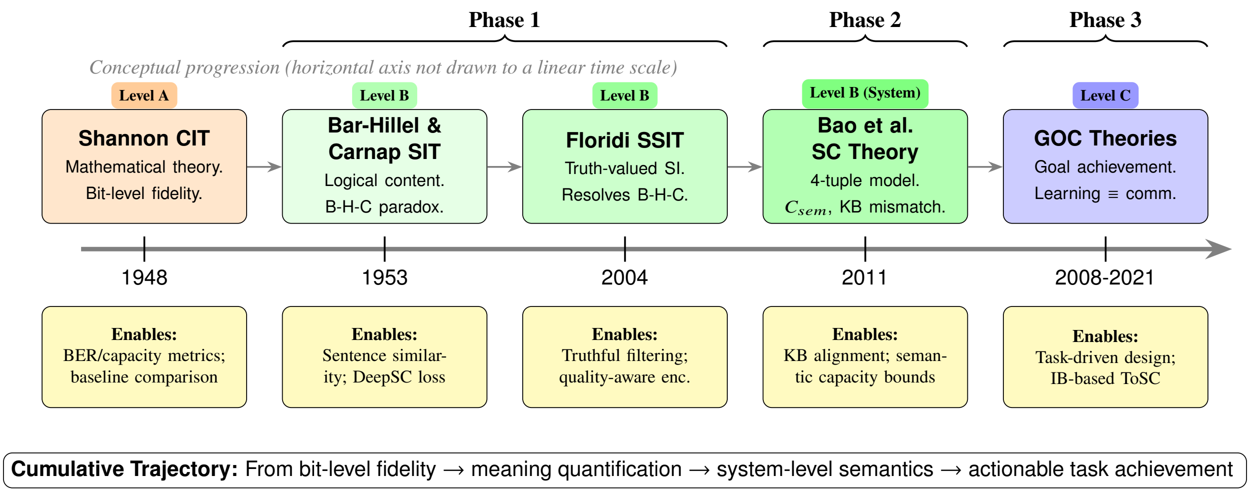

Figure 3: Theoretical evolution of semantic communication foundations. Five milestones trace the progression from Shannon's classical information theory (Level A, 1948) through semantic information quantification (SIT, SSIT; Level B, 1953–2004) and system-level semantic exchange modeling (Bao et al.; Level B, 2011), to goal-oriented communication theories (Level C, 2008–2021). Each forward-link box indicates how the historical contribution enables modern ToSC practice. Phase braces correspond to the three subsections of Section 2. The horizontal axis orders milestones by theoretical progression and is *not* drawn to a linear time scale; the year labels denote approximate periods, and goal-oriented communication theories (2008–2021) overlap the 2011 system-level work.

insight, that information content should be measured by its *semantic* rather than statistical properties, directly underpins modern SC system design: the sentence-similarity metrics used to train and evaluate systems such as DeepSC [37] are practical descendants of SIT's attempt to quantify meaning, replacing formal logic with learned neural embeddings (see Section 4).

2.2. Phase 2: Modeling End-to-End Semantic Exchange

Bao et al. [38] moved beyond individual-statement semantics to model SC as a system. Their framework characterizes both source and destination by a 4-tuple $\langle W, K, I, M\rangle$ (world model, background knowledge, inference engine, and message interpreter) and defines semantic channel capacity as follows:

$$C_{\text{sem}} = \sup_{P(X|W)} \left[I(X;Y) - H(W|X) + \overline{H}_s(Y) \right] \tag{1}$$

where $I(X;Y)$ is the Shannon mutual information between the sent and received messages (the syntactic channel term, as in CIT); $H(W|X)$ is the semantic-encoder *equivocation*, the residual ambiguity of the world model W given the transmitted message under the sender's knowledge base and inference procedure; and $\overline{H}_s(Y)$ is the average model-theoretic *semantic* information the receiver can recover from Y under its own knowledge base and inference procedure, distinct from the Shannon output entropy $H(Y)$. Because this interpretive term enters additively, Bao et al.'s Semantic Channel Coding Theorem yields a capacity that may lie *above or below* the classical capacity $\sup_{P(X|W)} I(X;Y)$, according to whether the receiver's semantic-interpretation gain $\overline{H}_s(Y)$ or the encoder equivocation $H(W|X)$ dominates. This capacity depends on the alignment between the source's and destination's knowledge bases (K_s vs. K_d): when $K_s \neq K_d$, semantic noise arises that has no counterpart in classical channel noise. The insight that *knowledge mismatch degrades semantic capacity independently of physical channel quality* is foundational for understanding the knowledge-base dependency that pervades modern ToSC systems (discussed in Section 3). However, the framework remained conceptual, relying on idealized assumptions (e.g., $K_s = K_d$) that are rarely satisfied in practice; moreover, the world model W is assumed to be finite and enumerable, which is unrealistic for continuous sensor data in applications such as autonomous driving or medical imaging. Güler et al. [39] subsequently formulated meaning recovery over a noisy channel as a Bayesian game of incomplete information, in which an external agent of unknown (adversarial or helpful) intent shapes the receiver's interpretation, establishing game-theoretic foundations for robust semantic exchange. This Phase 2 insight, that knowledge-base alignment is a system-level design parameter as important as channel quality, manifests directly in modern ToSC architectures: the knowledge-base synchronization mechanisms surveyed in Section 3 (e.g., Farshbafan et al.'s curriculum learning [40], Zhou et al.'s knowledge graph mapping [41]) are practical attempts to satisfy the $K_s = K_d$ assumption that Bao et al.'s capacity formula requires.

2.3. Phase 3: The Pragmatic Turn: Goal-Oriented Communication (Level C)

The transition to Level C required a fundamental reconceptualization: communication as a means to achieve goals rather than to transfer meaning. Juba and Sudan [42] demonstrated that two agents without a shared language can still communicate successfully if the receiver can verify whether its goal has been achieved, a result limited to problems in the PSPACE complexity class but conceptually significant. Their framework was subsequently generalized into Goal-Oriented Communication (GOC) theory [43, 44], which distinguishes *meta-goals* (internal intentions) from *syntactic goals* (observable outcomes) and proves that goal achievement is possible even without shared linguistic structure, provided agents are cooperative and can sense progress toward their objectives. A key theoretical result established that constructing universal communication strategies under this framework is equivalent to designing online learning algorithms [45], forging a direct link between communication theory and machine learning.

Kountouris and Pappas [46] further bridged these theoretical phases by introducing a semantics-empowered framework for networked intelligent systems that integrates timing, context, and goal-awareness into communication design. This progression, from SIT's logical semantics through system-level SC models to GOC's task-driven formulation, provides the theoretical justification for ToSC as a design philosophy. Where classical SC optimizes for semantic fidelity at Level B, ToSC inherits GOC's principle that communication success is measured solely by whether the receiver's task is accomplished.

However, these foundational works largely relied on trial-and-error communication paradigms and offered guidance for designing protocols or strategies for simple communication systems, such as server-printer interactions; their focus remained on a mathematical theory of GOC for traditional computer models. Despite their limited scope, these early GOC systems marked a milestone: the ultimate measure of communication success is task achievement, not symbol accuracy. Modern ToSC systems operationalize this principle through learned goal verification. For example, Kang et al.'s personalized saliency framework [13] lets the receiver's task define feature relevance, while Shao et al.'s IB-based edge inference [47] jointly trains encoder and task head, directly realizing Juba's theoretical link between learning algorithms and communication strategies.

3. ToSC Framework: A Specialization of SC

3.1. Defining ToSC

Traditional communication systems, grounded in Shannon's theory, prioritize reliable bit transmission while often neglecting semantic content or task-specific data utility, an approach increasingly insufficient for modern applications that prioritize informed machine decisions over mere data relay, such as autonomous driving, healthcare monitoring, and industrial automation. ToSC addresses these shortcomings by transmitting only task-relevant semantic content rather than all available data, acknowledging the varying value of information across applications and tailoring data to tasks for more effective resource allocation and communication efficiency [15].

ToSC has two primary objectives: effectiveness and resource conservation. Effectiveness ensures task success, defined by task-specific outcomes such as high classification accuracy or reliable robotic control, with emphasis on the receiver's functional outcome. The second, often the primary driver of ToSC adoption, is communication efficiency: reducing transmitted data by conveying only task-relevant semantics lowers bandwidth, latency, and potentially energy consumption, important for resource-constrained environments such as IoT and mobile settings [48].

Distinguishing ToSC from general SC clarifies its unique focus, despite their close relationship: Fig. 4 and Table 2 illustrate the workflow and paradigm differences. General SC aims to transmit meaning accurately and efficiently so the receiver's interpretation aligns with the sender's intent, whereas ToSC is a pragmatic specialization emphasizing task-specific utility: to enhance efficiency, it may intentionally discard semantically valid but task-irrelevant meaning that general SC, designed to preserve meaning faithfully, would retain [49]. For example, an image classification ToSC system may transmit only features discriminative for the target classes, ignoring background or texture details that a general semantic system focused on reconstruction may preserve. Goal-oriented communication, central to ToSC, inherently involves both understanding meaning and achieving a specific objective, so the receiver's primary concern is effectiveness in accomplishing the predefined goal rather than semantic completeness alone [14].

A defining characteristic of ToSC is its intrinsic reliance on shared knowledge accessible to both the transmitter and receiver [38, 7]. The term "knowledge base" has been applied loosely in the literature to everything from ontologies to pre-trained model weights, which obscures important distinctions. We therefore separate three layers of shared knowledge and reserve the term *knowledge base* for the first two:

- **Static knowledge base (KB):** fixed, explicitly shared, and addressable knowledge agreed upon a priori, such as ontologies, codebooks, label sets, and shared datasets or embeddings.
- **Dynamic context:** session- or runtime-state synchronized between endpoints during operation, such as recent observations, dialogue history, or an evolving task state.
- **Implicit (parametric) knowledge:** knowledge embedded in model weights (pre-trained or fine-tuned encoders, and large language and multimodal models). We treat this as a *carrier* of knowledge rather than a knowledge base in itself: a model is a source of implicit knowledge that can *back* or *substitute for* an explicit KB, for instance through retrieval-augmented generation, but it is not directly addressable or jointly editable in the way a shared KB is.

Across these layers, shared knowledge (1) establishes the context that makes "task-relevant semantics" well-defined, (2) reduces transmission volume by letting communication focus on information not already inferrable from shared knowledge, and (3) enables intelligent encoding and decoding at both ends of the link [20]. We use this three-layer terminology consistently throughout the survey; the interplay between implicit parametric knowledge and explicit KBs in large-model-based systems is examined in Section 6.

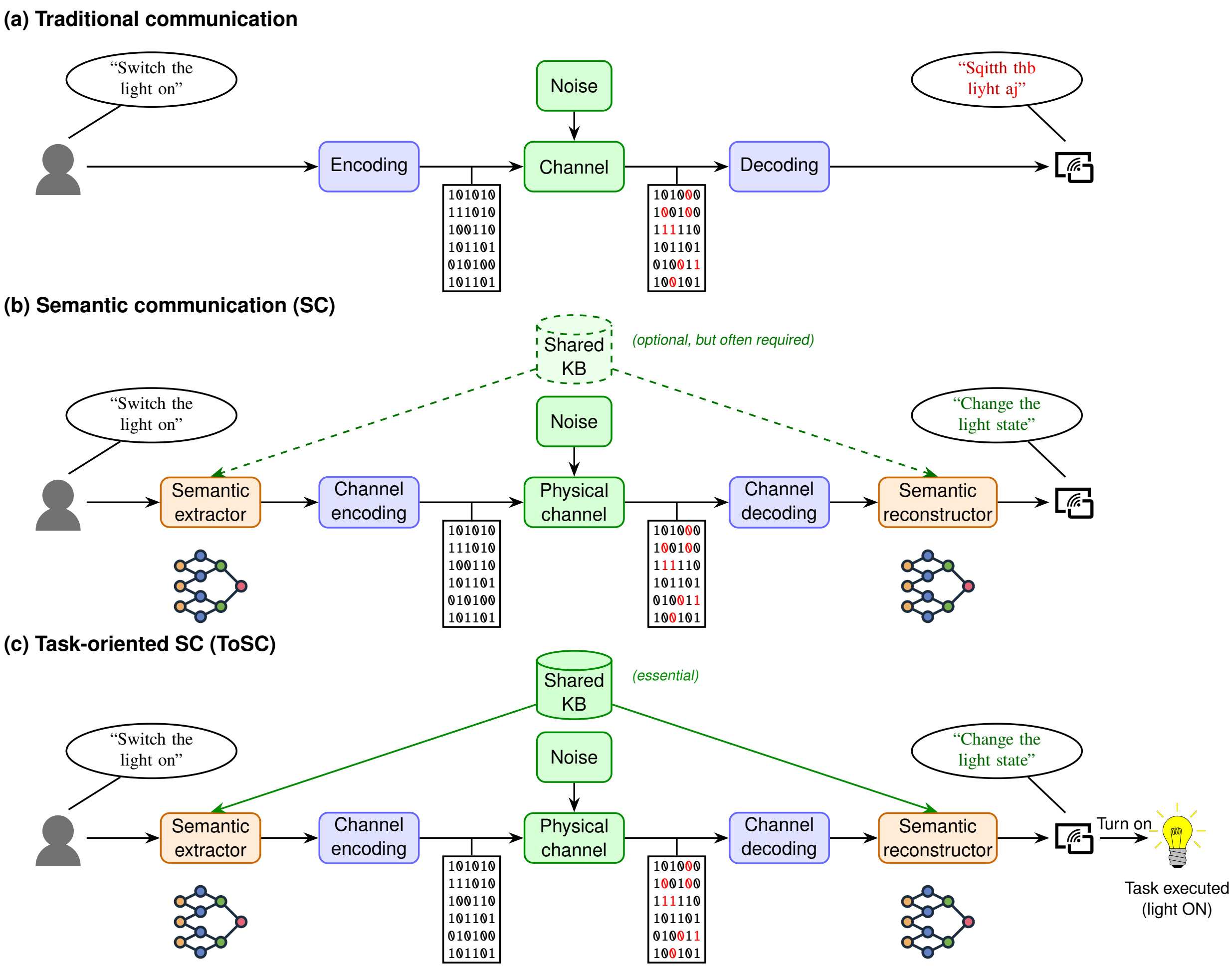

Figure 4: Comparative study of traditional models, SC systems, and ToSC systems: workflow of a) traditional communication, b) SC, and c) ToSC. In a) every bit must be recovered, so channel errors corrupt the delivered message. In b) the semantic extractor and reconstructor exchange a learned semantic representation, and the receiving system recovers the intended meaning despite bit errors. In c) the ToSC pipeline conveys only the semantics the receiving system needs to execute the task (switching the light on), rather than the full semantic content. A shared knowledge base supports semantic interpretation in both paradigms; as marked in the figure, it is optional, but often required, for general SC (dashed) and essential for ToSC's task relevance and interpretation (solid; Table 2).

Table 2: Comparison of Communication Paradigms

Feature	**Traditional (Shannon)**	**General Semantic Communication**	**ToSC**
Primary Goal	Accurate symbol/bit recovery	Faithful conveyance of meaning (semantics)	Effective task completion / Goal achievement
Focus Level	Technical (Level A)	Semantic (Level B)	Effectiveness (Level C)
Key Metrics	BER, SER, Capacity, Throughput	Semantic Similarity, Semantic Error	Task Success Rate, Task Error, Efficiency
Info. Value	All bits treated equally	Based on contribution to overall meaning	Based on relevance to specific task/goal
Role of Task	Agnostic	Implicitly related to meaning extraction	Central driver of communication design
Knowledge Base	Generally absent	Often required for semantic interpretation	Essential for task relevance & interpretation

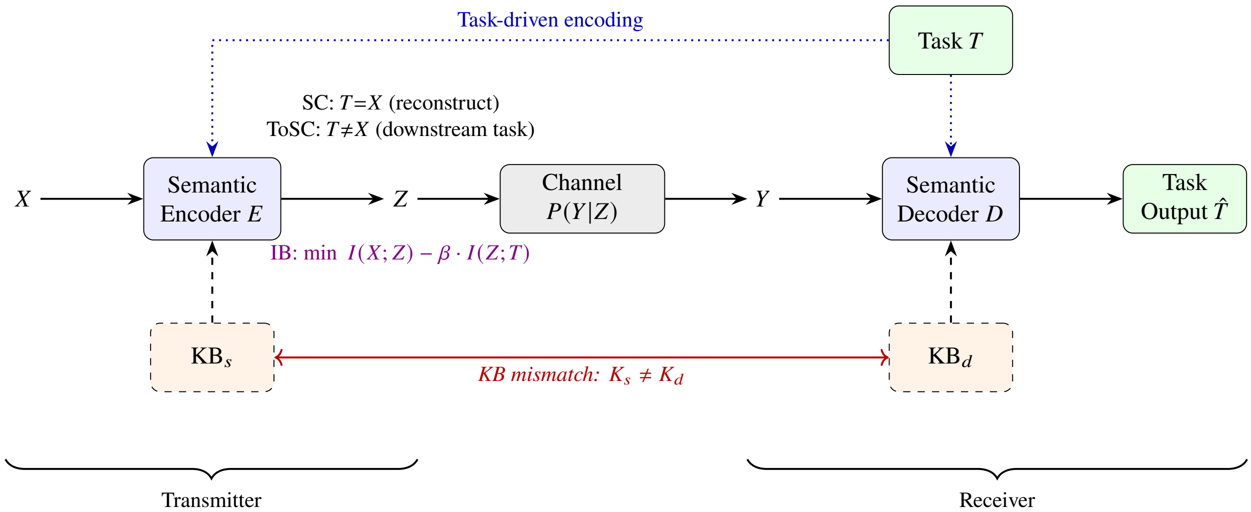

Figure 5: Consolidated ToSC system model with IB notation. The semantic encoder E, conditioned on source knowledge base KB_s, compresses source X into representation Z. The decoder D, conditioned on KB_d, produces the task output $\hat{T}$. The IB objective governs the compression–task-relevance trade-off: SC corresponds to T=X (full reconstruction), while ToSC targets a downstream task T≠X, discarding task-irrelevant information. The red link highlights that KB mismatch ($K_s \neq K_d$) introduces semantic noise independent of channel quality.

3.2. Functional Components and System Building Blocks

ToSC systems require a distinct set of functional blocks compared to traditional architectures, reflecting their emphasis on meaning and task execution [15, 50]. Fig. 5 presents a consolidated system model with IB notation. While specific implementations may vary, several core components are typically present.

3.2.1. Task-Tailored Semantic Encoders and Decoders

At the core of a ToSC framework are the semantic encoder and decoder, which replace or augment the conventional source and channel coding blocks. The semantic encoder analyzes the source data, often using contextual information or a KB, to extract task-relevant features and prepare them for transmission, typically as a compressed representation such as a latent vector.

The semantic decoder receives the (potentially noisy) transmitted information and interprets it, often using shared knowledge or contextual cues; its function depends heavily on the task, ranging from directly performing the task (e.g., classifying the input based on the received semantic features) to reconstructing a representation of the source data that contains the information required by a downstream task execution unit.

These encoders and decoders are task-tailored, with design and training optimized specifically for the intended goal. Deep neural networks (DNNs) commonly implement these components for their ability to learn complex mappings and extract salient features from data, for example convolutional autoencoders (CAEs) for image-related tasks [51], convolutional neural networks (CNNs) for feature extraction and classification [52, 53], and Transformer architectures for sequential or relational data such as text or multimodal inputs [37, 54]. A key feature of several modern ToSC systems is JSCC and end-to-end training, where the semantic encoder, channel transmission (or its model), semantic decoder, and task execution module are jointly optimized as a single system to maximize task performance under given communication constraints [51, 55]; this end-to-end deep JSCC formulation is examined in detail in Section 3.3.

3.2.2. Role of KBs and Context Acquisition

Unlike traditional, content-agnostic systems, ToSC systems frequently rely on shared knowledge or context between transmitter and receiver: this understanding, often formalized as a KB, is essential for both extracting relevant semantics at the transmitter and correctly interpreting them at the receiver [4]. The KB can take various forms depending on the application, including background information, ontologies defining terms and relationships, shared datasets used to train the semantic codecs, base models for rendering (such as an avatar's initial pose and mesh in AR), or task models trained in the cloud and broadcast to all users, as in multi-user VQA transceivers [12], and SC effectiveness depends heavily on this shared knowledge's quality and consistency.

In addition to static KBs, dynamic context acquisition modules can also play an important role [56, 46]: by sensing the environment, tracking user intent, or maintaining a history of interactions (potentially via memory modules), they let the system dynamically adapt its semantic processing to the current situation. For example, recent architectures that integrate large models let the foundation model's embedded knowledge stand in for a dedicated shared knowledge base, segmenting content by task prompt and reconstructing it with a conditional generative model, thereby operating without explicit KB synchronization [57].

However, the reliance on KBs is itself a liability: a KB must be built, kept up to date, and kept consistent across parties, and each requirement becomes harder in dynamic multi-agent systems. Semantic mismatch, where the KBs at the transmitter and receiver differ or become outdated, can lead to misinterpretation and ultimately task failure [38]. Recent approaches address this through curriculum learning frameworks that actively synchronize a common language across communication links [40], and through mapping abstract KBs to computable knowledge graphs that provide reproducible mechanisms for resolving semantic ambiguity [41]. Addressing these KB management issues, including modeling knowledge evolution, is therefore critical for the widespread deployment of reliable ToSC systems [20].

3.2.3. Task Execution and Evaluation Units

Located on the receiver side, the task execution unit uses the decoded SI to perform the intended goal [12, 11]; its nature is entirely application-dependent: (1) a classifier that outputs a label, (2) a controller that generates actuation commands for a robot, (3) a decision-making algorithm, (4) a rendering engine that reconstructs a scene, or (5) an inference engine that answers a question.

Complementing the execution unit, an evaluation mechanism, either implicit or explicit, assesses the task execution unit's performance against predefined metrics, important for system optimization during training (e.g., calculating the loss function based on task error) and potentially for adapting system behavior during operation. The performance metrics employed are detailed in Section 4.

To illustrate the ToSC framework, we consider a VQA system where the task is answering a textual question about an image (the source data). Here, the task-tailored semantic encoder typically employs a CNN for visual feature extraction and a Transformer for interpreting the question's semantics, fusing both into a joint, task-relevant semantic representation. This encoding and decoding process draws on shared knowledge: in VQA, a static KB of common-sense ontologies about objects and relationships, and contextual data from training sets, complemented by the implicit knowledge carried in the pre-trained encoder weights. If the VQA system components are distributed (edge processing versus a central server), the semantic representation is transmitted over a communication channel; the semantic decoder receives it and, often using multimodal reasoning techniques, interprets the fused semantics to extract the information needed to answer the question; the task execution unit renders this information as the final textual answer. An evaluation unit then assesses success using metrics such as VQA accuracy, comparing the generated answer to the ground truth.

3.2.4. Interaction between Components

This end-to-end flow, traced concretely in the VQA example above, is common to ToSC systems regardless of application: guided by the task requirements, the KBs, and contextual information, the encoder's compressed representation crosses the channel to a semantic decoder that either directly produces the task output or hands an intermediate representation to the task execution unit for execution and evaluation.

ToSC frameworks are defined by tight integration and often joint optimization of these functional blocks [15], which differs from the traditional modular approach, where source coding, channel coding, and modulation are optimized independently; this separation's optimality guarantees do not extend to the finite-length regimes typical of ToSC [58]. Architectural variations exist, such as split learning, where the encoder resides on an edge device while the decoder and task execution unit operate on a more powerful server for distributed resource usage.

3.3. Implementation Architectures

The conceptual framework of ToSC has been realized through diverse architectural designs. We structure the architectural taxonomy along four functionally meaningful axes that directly inform engineering decisions: (1) encoding strategy (end-to-end JSCC vs. modular separate source-channel coding), (2) transmission mode (analog vs. digital), (3) task-relevance mechanism (how the system identifies and prioritizes task-relevant information), and (4) adaptivity (static vs. channel-adaptive vs. task-adaptive). These axes capture the semantic communication *design logic* that the neural network backbone alone does not; the AI model family is complementary, determining *how* each axis is realized, and is analyzed separately in Section 3.4. Table 3 maps representative systems along the four axes. These dimensions are not individually new: the joint-vs-separate and analog-vs-digital distinctions are established JSCC design choices (cf. the overview of Gündüz et al. [15]), and task-relevance and adaptivity recur throughout the ToSC literature; our contribution is consolidating them into a single comparative frame overlaid with the model-family lens of Section 3.4, not introducing each axis individually.

3.3.1. Encoding Strategy: End-to-End JSCC vs. Modular

The most fundamental architectural distinction in SC/ToSC is whether source and channel coding are performed jointly or separately.

End-to-end JSCC systems train a single DNN that maps source data directly to channel input symbols, bypassing the traditional separation of source coding, channel coding, and modulation. Bourtsoulatze et al. [51] demonstrated that such deep JSCC encoders achieve better rate-distortion performance (measured by PSNR and SSIM under noisy channels) than separated source-channel coding, establishing the foundational architecture for modern systems. Subsequent work expanded this: Kurka and Gündüz [59] introduced bandwidth-adaptive DeepJSCC with successive refinement, transmitting images progressively in layers that receivers aggregate to trade bandwidth for reconstruction quality with graceful degradation. Dai et al. [60] proposed NTSCC, which bridges neural transform coding and JSCC by learning a latent entropy model that guides joint rate-distortion-channel optimization, achieving gains over both purely learned compression and conventional JSCC. Kurka and Gündüz [61] further improved wireless image transmission by exploiting channel output feedback to enable iterative refinement at the transmitter (DeepJSCC-f); for a detailed treatment of deep JSCC, see the overview by Gündüz et al. [15]. This approach offers graceful degradation under varying channel conditions: performance degrades smoothly rather than exhibiting the cliff effect of digital systems, and naturally supports variable-rate transmission when combined with bandwidth-adaptive mechanisms. However, end-to-end JSCC requires differentiable channel models for gradient-based training, which limits applicability to channels with tractable mathematical descriptions.

Modular (separate coding) systems retain the classical separation principle, applying a learned semantic encoder for source compression followed by conventional or learned channel coding. This approach offers compatibility with existing communication infrastructure and allows independent optimization of each module, exemplified by systems such as the VQA framework of Xie et al. [12], which extracts visual semantic features with a frozen ImageNet-pretrained ResNet-101 encoder before standard transmission. The drawback is that the separation theorem's optimality guarantees, which hold for infinite block lengths, do not extend to the finite-length, delay-constrained regimes typical of ToSC applications, where JSCC can offer substantial gains [15].

3.3.2. Transmission Mode: Analog vs. Digital

Within JSCC, a secondary but consequential distinction is whether the encoder output is transmitted as continuous-valued (analog) channel symbols or quantized to a finite constellation (digital).

Analog JSCC transmits continuous-valued encoder outputs directly over the channel, preserving the smooth mapping between source semantics and channel symbols. This enables graceful degradation: as the signal-to-noise ratio (SNR) varies, reconstruction or task quality degrades proportionally rather than catastrophically. Most deep JSCC systems, including Bourtsoulatze et al. [51] and the multi-task framework of Lyu et al. [52], operate in this mode. Analog operation, however, is incompatible with digital communication infrastructure (e.g., existing modulation and coding schemes, ARQ protocols).

Digital JSCC quantizes encoder outputs to discrete symbols compatible with standard digital modulation, enabling integration with existing protocol stacks and standard error correction. Tung et al. [62] proposed DeepJSCC-Q, which learns to map continuous encoder outputs to a finite constellation while preserving the graceful degradation property through a soft-to-hard quantization strategy, demonstrating that digital JSCC can approach the performance of its analog counterpart. Quantization noise is the cost, and hybrid approaches using learned quantization to minimize the quantization-task performance gap remain an active research direction.

3.3.3. *Task-Relevance Mechanism*

A defining design choice in ToSC is how the system identifies which information is relevant to the downstream task, an axis along which three approaches are prevalent.

Implicit (loss-function-driven): The most common approach defines task relevance implicitly through end-to-end training with a task-specific loss function (e.g., cross-entropy for classification, mean average precision (mAP) for detection). The encoder learns to retain task-relevant features and suppress irrelevant ones as a byproduct of optimizing the loss. Encoder-decoder pairs trained end-to-end for a classification task [63] and autoencoder-based systems trained with reconstruction loss exemplify this approach. Its appeal is simplicity, but relevance is fixed at training time and cannot adapt to new tasks without retraining.

Explicit (ranking/filtering): Some systems decouple feature extraction from relevance assessment by applying a separate ranking or filtering stage. The personalized encoder of Kang et al. [13] scores scene-graph triplets by their saliency to a user's stated interest and transmits only the matching subgraph, and the goal-oriented avatar framework of Wang et al. [64] transmits skeleton keypoints ranked by their importance to avatar pose recovery. This approach supports dynamic relevance adaptation but requires an explicit relevance model.

IB-principled: A theoretically grounded alternative uses the information bottleneck framework (Eq. 2) to derive representations that are maximally informative about the task while maximally compressing the source. The deep variational information bottleneck [65] provides tractable bounds for DNN-based optimization. The IB approach offers a principled compression-relevance trade-off controlled by a single parameter β, but is computationally intensive and sensitive to the accuracy of variational approximations.

3.3.4. *Adaptivity*

ToSC systems vary in their ability to adapt to changing conditions during deployment.

Static: The encoder and decoder are fixed after training, with performance optimal for the training distribution but degrading under distribution shift (channel variation, task drift, KB mismatch). Most early ToSC systems operate in this mode.

Channel-adaptive: The encoder adjusts its output based on real-time channel state information (CSI). The channel-adaptive gated network of Lyu et al. [52] prunes output features based on channel SNR, dynamically adjusting the data rate while maintaining task performance. Chen et al. [66] use a masked auto-encoder with a channel-aware extractor that selects patches for transmission based on CSI.

Task-adaptive: The encoder adjusts its behavior based on changing task requirements. Multi-task learning architectures [53] learn shared representations that serve multiple tasks via task-specific decoder heads, though keeping the tasks' features separable within one backbone requires dedicated remedies such as the domain-adaptation design of Zhang et al. [67], which projects each task's features into a task-specific domain so that only those features need to be transmitted. More flexible approaches use foundation models (LLMs/LMMs) to dynamically decompose and adapt to new tasks [68], though at substantial computational cost.

3.3.5. *Advanced Integration Approaches*

Beyond the core design axes, recent architectures extend ToSC through integration with emerging technologies: O-RAN integration embeds semantic intelligence within the radio access network via a semantic plane and semantic RIC [56], and the broader design space for embedding such semantic processing across the radio-access plane is surveyed by Meng et al. [23]. Hybrid approaches also combine ToSC with complementary paradigms such as Integrated Sensing and Communications (ISAC), where communication signals serve dual sensing purposes [48]; these are discussed in detail as research trends in Section 6.

3.4. *The Role of AI Model Families in ToSC*

The functional axes above describe *what* architectural choices a ToSC system makes; the AI model family determines *how* those choices are realized, and is where the engineering of semantic extraction and task inference actually happens. Because these methods are central to the artificial-intelligence perspective on ToSC, we analyze the model families that recur across the surveyed systems along four roles (semantic extraction, semantic compression, task reasoning, and knowledge representation) and summarize their suitability and limitations relative to the architectures of Section 3.3 in Table 4.

Table 3: Taxonomy of Representative ToSC Architectures Along Functional Design Axes

Representative System	Encoding Strategy	Tx Mode	Task-Relevance	Adaptivity	Backbone	Domain / Task
Bourtsoulatze et al. [51]	JSCC (end-to-end)	Analog	Implicit (recon. loss)	Static	CAE	Image reconstruction
Lyu et al. [52]	JSCC (end-to-end)	Analog	Implicit (unified loss)	Channel-adaptive	CAE + gated net	Multi-task (recon. + classif.)
Eldeeb et al. [53]	JSCC (end-to-end)	Analog	Implicit (MTL loss)	Task-adaptive	CAE/CNN + MTL heads	Autonomous vehicles
Xie et al. [12] (MU-DeepSC)	Modular	Digital	Implicit (task loss)	Static	ResNet-101 + MAC	VQA (multi-user)
Wang et al. [69] (TSAR)	Modular	Digital	Explicit (base-knowledge selection)	Static	SANet	AR avatar rendering
Chen et al. [66]	JSCC (end-to-end)	Analog	Explicit (patch scoring)	Channel-adaptive	Masked AE	Flexible multi-task
Wang et al. [70] (I-JSCC)	JSCC (end-to-end)	Analog	Explicit (feature importance)	Channel-adaptive	CNN	Multi-user image
Ma et al. [71]	Modular	Digital	IB-principled (β-VAE)	Static	β-VAE	Explainable SC
Du et al. [72]	Modular (split)	Digital	Implicit (LMM reasoning)	Task-adaptive	LMM + lightweight enc.	Vehicular VQA
Strinati et al. [56]	Modular (O-RAN)	Digital	Explicit (semantic plane)	Task-adaptive	Protocol-level	Network-wide mgmt.

Autoencoders and variational autoencoders form the backbone of deep JSCC: a learned encoder–bottleneck–decoder maps the source directly to channel symbols, performing extraction and compression jointly [51]. Disentangled variants such as the β-VAE expose interpretable latent factors that can be selectively transmitted [71]. They excel at analog end-to-end JSCC but offer little explicit task reasoning, and their latent space is largely opaque, which complicates structured knowledge representation and interpretability.

Convolutional neural networks are the dominant choice for spatial and visual semantic extraction, encoding images into compact, task-discriminative features for classification- or detection-oriented ToSC [52]. Their inductive bias and hardware efficiency suit channel-adaptive, loss-driven (implicit-relevance) designs, but they capture local structure better than long-range or relational dependencies and do not natively represent structured knowledge.

Transformers and Vision Transformers model long-range dependencies and support cross-modal fusion through attention, underpinning text SC (e.g., DeepSC [37]) and multimodal, multi-user systems [54]. Attention also provides a natural mechanism for explicit task-relevance ranking. The cost is computational: quadratic attention and large parameter counts raise latency and energy use, which constrains deployment on resource-limited edge devices.

Graph neural networks (GNNs) operate directly on graph-structured semantics such as scene graphs and knowledge graphs, making them the natural substrate for knowledge representation and relational task reasoning; they underpin explainable scene-graph SC [73] and cross-modal graph-based SC [74]. The dependence on an explicit graph-construction step is the main weakness: when the extracted relations are noisy, incomplete, or absent, the reasoning built on them is degraded accordingly.

Large language and multimodal models act simultaneously as cross-modal semantic codecs, task decomposers, and carriers of the implicit (parametric) knowledge identified in our three-layer knowledge definition; the KG-LLM scheme of Salehi et al. [75] illustrates the coupling of an explicit graph with parametric knowledge. They offer the strongest task reasoning and zero-shot adaptivity of the families surveyed, but heavy inference cost, hallucination, and inconsistency across model versions and fine-tunes restrict real-time and edge use, as discussed further in Section 6. The moderate semantic-compression rating in Table 4 reflects the model family's intrinsic role as a within-modality codec, no denser than a tuned autoencoder bottleneck; the order-of-magnitude reductions associated with large-model

ToSC (e.g., the extreme image-to-text compression of Ribouh et al. in Section 6.1.1) arise instead from the cross-modal transform the model family enables (representing an image by a short text description), a distinct mechanism from the bottleneck compression scored in that column.

No single family dominates all four roles, and practical ToSC systems compose them, for example a CNN or ViT extractor feeding a GNN- or LLM-based reasoner, followed by an autoencoder bottleneck for transmission. The functional-axis taxonomy and this model-family lens are therefore complementary: the axes fix the communication design logic, while the model family determines the achievable semantic granularity, reasoning depth, and deployment cost.

Table 4: AI Model Families in ToSC: Suitability Across Semantic Roles

Model Family	**Semantic Extraction**	**Semantic Compression**	**Task Reasoning**	**Knowledge Repr.**	**Representative Use / Key Limitation in ToSC**
Autoencoder / VAE	•	•	×	×	Backbone of deep JSCC [51]; β-VAE for interpretable selective transmission [71]. Opaque latent space; little task reasoning.
CNN	•	◦	×	×	Spatial/visual feature extraction for classification-oriented ToSC [52, 53]. Weak at long-range and relational reasoning.
Transformer / ViT	•	◦	•	◦	Long-range dependencies and cross-modal fusion; text SC [37] and multimodal multi-user SC [54]. High compute and latency.
GNN	◦	◦	•	•	Graph-structured semantics (scene/knowledge graphs); explainable scene-graph SC [73] and cross-modal graph SC [74]. Needs explicit graph construction; brittle to noisy relations.
LLM / LMM	•	◦	•	•	Cross-modal codec, task decomposition, and carrier of implicit knowledge [75]. Inference cost, hallucination, version/fine-tune drift.

Suitability per role (operational scoring): • = a primary, purpose-built mechanism for the role in the surveyed ToSC literature; ◦ = capable of the role but as a secondary function or with known limitations; × = not an inherent strength, typically delegated to another family. Generative and diffusion models are not listed as a separate family because they act as a *decoding* mode rather than an extraction/compression backbone; they are treated in Section 6.1.4. *Mapping to the functional axes of Table 3*: autoencoders/VAEs realize analog end-to-end JSCC and, in disentangled β-VAE form, the information-bottleneck-principled modular design of Ma et al. [71]; CNNs suit channel-adaptive, implicit-relevance designs; Transformers/ViTs enable explicit task-relevance ranking; GNNs underpin knowledge-structured, relational modular designs such as the structured scene-graph SC of Sun et al. [73]; and LLMs/LMMs enable task-adaptive modular reasoning. The Backbone column of Table 3 gives the realized per-system pairing. These per-role suitability ratings reflect the authors' assessment of the surveyed literature rather than a numerical benchmark, and a single rating necessarily abstracts over variation within each broad family.

3.5. Semantic Processing Mechanisms

ToSC systems process information through three stages that collectively determine what is transmitted: representation, extraction, and filtering. Task-relevant semantics are represented as latent vectors (learned by AE/DNN encoders) [76], knowledge graphs (subject-predicate-object triples for structured data) [73, 41], or explicit feature vectors (bounding boxes, skeleton joints, classifier inputs) [77, 64]. Relevance is defined either implicitly [63] or explicitly [77, 64], per the task-relevance axis of Section 3.3.

Extraction techniques include (1) DNN-based feature learning (CNNs, Transformers, AEs), (2) classical CV algorithms (object detectors, keypoint estimators) [8], (3) NLP methods (end-to-end text semantic encoders and knowledge-graph construction) [78, 41], and (4) the IB framework [26], which derives compressed representations maximizing task-relevant mutual information via variational approximations [65, 50]. After extraction, semantic ranking prioritizes elements by task contribution, adaptive transmission adjusts fidelity to channel conditions and target reconstruction quality [79], and semantic filtering removes task-irrelevant content [80].

This pipeline entails a fundamental trade-off. Blau and Michaeli [81] proved that rate-distortion bounds conflict with perceptual quality, motivating perceptual/generative over pixel-fidelity compression, a distinction within semantic fidelity (Level B). The case for task-oriented (Level C) compression rests instead on the information-bottleneck argument: Shao et al. [47] used a variational information-bottleneck formulation to characterize an achievable rate–accuracy trade-off between communication rate and task performance. Recent work refines this IB-based rate control in several directions: rate-optimized IB with SNR-adaptive networks for resilient image SC [82], feature disentanglement into common and private components to prune redundancy in multi-device cooperative ToSC [83], and a conditional rate-distortion formulation with task prompts and feedback for task-adaptive coding [84]. Defining relevance too narrowly risks brittleness; finding the optimal balance between compression and task effectiveness remains a central challenge.

3.5.1. Cross-Mechanism Synthesis

Some of the surveyed mechanisms recur far beyond the domains that introduced them. *Information-theoretic extraction* (IB-based or variational) and *saliency-based prioritization* appear in nearly every application domain, from text and image transmission to autonomous driving and extended reality, and form the foundational layer of any ToSC pipeline. *Adaptive transmission* and *feedback coupling* recur for a different reason: real-world channels and tasks are non-stationary, and static semantic compression breaks down once the environment or the task objective evolves. *Robustness control* mechanisms (adversarial training, channel-adaptive encoding, redundancy injection) appear wherever deployment conditions are uncertain, reflecting the tension between aggressive semantic compression and system reliability. Supplementary Table S2 maps each recurring mechanism to representative systems, application domains, and the underlying reason for its recurrence.

3.6. When Does the SC/ToSC Distinction Matter? A Critical Analysis

The preceding sections established that SC optimizes for semantic fidelity (Level B) while ToSC optimizes for task effectiveness (Level C). However, this distinction risks being merely definitional unless we can identify *when it leads to materially different system behavior*. This subsection synthesizes the IB framework, originating with Tishby et al. and mapped to SC by Beck et al., with empirical evidence from the surveyed literature to characterize the qualitative regimes under which SC and ToSC diverge or converge, together with the design principles that follow.

3.6.1. Formal Grounding via the Information Bottleneck

The SC/ToSC distinction can be formalized through the IB framework [26, 85]. Beck et al. [27] cast this IB formulation in a wireless physical-layer semantic-communication setting, using a mutual-information objective to trade off source-information preservation against task-relevance. Given source data X, a compressed representation Z, and a task variable T, the IB objective minimizes:

$$\mathcal{L}_{\mathrm{IB}} = I(X;Z) - \beta \cdot I(Z;T) \tag{2}$$

where $I(\cdot;\cdot)$ denotes mutual information and β controls the trade-off between compression and task-relevance. General SC corresponds to the regime where $T = X$ (the task is full reconstruction), so $I(Z;T) = I(Z;X)$, and the system preserves as much source information as possible at a given rate. ToSC corresponds to $T \neq X$ (the task is a downstream function of the source, e.g., a class label), and the system discards source information that is irrelevant to T.

This formalization reveals that **SC and ToSC diverge whenever** $I(X;T) < H(X)$, that is, whenever the task requires strictly less information than the source contains. The gap $H(X) - I(X;T)$ quantifies the "task-irrelevant" information that ToSC can discard but SC must preserve. For high-dimensional sources (images, video, point clouds) with low-dimensional tasks (classification, detection), this gap is typically large, making the distinction practically significant. The strict inequality itself holds for essentially any lossy task (since $H(X) - I(X;T) = H(X|T) \geq 0$, with equality only when X is a deterministic function of T); it is therefore the *magnitude* of this conditional entropy, not the inequality alone, that decides whether the SC/ToSC distinction is of engineering significance.

3.6.2. Concrete Evidence of Divergence

The surveyed literature, complemented by a representative vehicular design pattern, shows that SC-optimized and ToSC-optimized systems produce materially different outcomes:

Case 1: Personalized task-oriented selection vs. full reconstruction. Kang et al. [13] design a personalized, saliency-based task-oriented encoder for UAV image sensing that represents each image as a scene graph of subject-relation-object triplets and transmits only the triplets matching a user's stated interest (e.g., "man wearing a jacket"), weighted by a personalized attention map, reporting on the order of a 64% reduction in communication cost. A reconstruction-oriented SC system would instead preserve the entire image; the two systems thus retain materially different content from the same source, illustrating how a task objective and a reconstruction objective *diverge in what they preserve*. Xiao et al. [86] formalize a related divergence through rate-distortion theory for strategic semantic communication, in which transmitter and receiver hold different distortion measures and choose their encoding and decoding strategies rationally, showing that the achievable rate-distortion performance depends on the equilibrium the two sides reach rather than on signal fidelity alone.

Case 2: Autonomous driving: detection vs. scene reconstruction. As a representative vehicular-network design pattern, a ToSC system transmitting only detected object features (bounding boxes, class labels, trajectories) would achieve high mAP for downstream planning using a fraction of the bandwidth required for full scene reconstruction, but such systems fail when the downstream task changes unexpectedly, for example when a previously unmodeled obstacle type appears, or when the planning module requires contextual information (road surface conditions, weather cues) discarded by the task-specific encoder. This reveals a fundamental **brittleness trade-off**: ToSC's efficiency gains come at the cost of reduced adaptability to task drift.

Case 3: Multi-task systems as the boundary case. Lyu et al. [52] design a multi-task SC system that jointly optimizes for image reconstruction (Level B) and classification (Level C). By sweeping the loss-balance parameter between the two objectives, they show that the two cannot be maximized simultaneously: as the weighting shifts toward classification, the classification accuracy rises while the reconstruction PSNR falls (their Fig. 10), tracing a non-trivial, qualitatively reported accuracy–PSNR trade-off between Level B and Level C performance. This trade-off is consistent with the information-bottleneck picture above. Because the classification target is a coarsening of the source (T a deterministic function of X), the data-processing inequality bounds task-relevant by reconstruction-relevant information, $I(Z;T) \leq I(Z;X)$. We caution that this is an information-content bound, not a guarantee that task-relevant *features* form a literal subset of reconstruction-prioritized ones. As Case 4 below notes, low-variance yet high-value features can be task-critical while a reconstruction objective deprioritizes them. The observed trade-off is therefore suggestive evidence for this bounded-information relationship in their multi-task setting rather than a proof of feature-set containment.

Case 4: Low-variance, high-value features. Divergence can also arise at the opposite extreme of compression, when task-critical information is low-variance yet high-value: a small lesion occupying a few pixels of a medical image, or a distant pedestrian in a driving scene. Variance-driven compression preserves the dominant background while discarding precisely the subtle, low-energy features on which the task depends; the target then occupies a small region within a large task-irrelevant background, so the effective task-information gap $H(X) - I(X;T)$ is *large* despite the target's low variance. Given a relevance signal, a task-aware encoder can protect these high-value features that a reconstruction-oriented objective would deprioritize, so SC and ToSC *diverge*. Without such a signal (as in rare-pathology cases where the salient feature is not flagged), both objectives discard the feature and task outcomes degrade in tandem, a *co-failure* rather than a benign convergence.

3.6.3. When the Distinction Collapses

The SC/ToSC distinction becomes negligible in two important scenarios:

(1) Low-rate regimes with energy-dominant tasks. When the channel bandwidth is severely limited *and the task's discriminative information is carried by the highest-variance (energy-dominant) components of the source*, both SC and ToSC systems are forced onto the same small set of most-salient features. Under this condition, and only under it, the features necessary for reconstruction and those necessary for the task converge, because both must prioritize the highest-variance components of the data. This is an expected consequence of the rate-constrained regime: at very low SNR or extreme compression ratios, both objectives are forced onto the same high-variance components, so SC- and ToSC-optimized systems should perform similarly. This *success-convergence* (both objectives retain the same features and the task still succeeds) must be distinguished from the *co-failure* of Case 4 above, in which both objectives discard a low-variance but task-critical feature and degrade in tandem: the former makes ToSC unnecessary, the latter makes it ineffective without a relevance signal. The convergence argument here therefore applies only when the task's discriminative information is energy-dominant, and should not be assumed for fine-grained or safety-critical perception.

(2) Tasks that require full source fidelity. Some tasks (open-ended diagnostic reading, where a clinician may inspect the entire image for any abnormality so that almost all of the source is potentially task-relevant, or speech communication where prosodic nuances carry meaning) inherently require high source fidelity. For these, the "task-irrelevant" information gap $H(X) - I(X;T)$ is small, and ToSC offers minimal advantage over SC. This observation has practical implications: *not all applications benefit from the ToSC approach*, and engineers should evaluate the information gap before investing in task-specific encoder design.

3.6.4. Implications for System Design

This analysis yields three practical design principles for practitioners:

1. **Estimate the task-information gap before committing to ToSC.** If the downstream task is complex (e.g., multi-step reasoning, open-ended generation) or if task requirements may change, the efficiency gains of ToSC may not justify the loss of generality. A multi-task or progressive encoding approach may be more appropriate.
2. **KB alignment is plausibly more critical for ToSC than for SC.** We argue that, because ToSC systems discard "redundant" source information a receiver could otherwise exploit to recover from KB mismatch, they should be *more sensitive* to KB degradation than general SC systems, which retain that information (cf. Section 3). We present this as a reasoned design conjecture rather than an empirically settled result: the surveyed literature establishes that KB mismatch degrades semantic capacity for SC broadly [38] but does not isolate a strictly greater task-performance degradation for ToSC than SC under matched KB drift. Quantifying it rigorously requires the KB-dependent rate-distortion bounds we flag as Priority 1 (Section 7). This conjecture nonetheless has direct implications for deployment in dynamic environments where KBs evolve independently.
3. **Design for graceful degradation.** ToSC encoders should include mechanisms for fallback to higher-fidelity transmission when task confidence is low or when the system detects distribution shift, rather than operating at a fixed compression point optimized for a single task scenario.

Each of these principles presupposes that semantic fidelity and task effectiveness can actually be measured; Section 4 examines the metrics available for doing so.

4. Performance Evaluation Metrics

Evaluating SC and ToSC systems requires metrics that go beyond classical bit error rate to capture meaning preservation and task success. Following the three-level framework established in Section 2 (cf. Fig. 2), this section organizes metrics by their target level: Level B (semantic fidelity) for general SC, Level C (task effectiveness) for ToSC, and cross-level efficiency metrics. We conclude with an analysis of metric conflicts and practitioner guidance for metric selection. Supplementary Table S3 provides a consolidated overview of metrics across all levels.

4.1. Metrics for Semantic Fidelity and Quality (Level B)

Level B metrics assess how faithfully meaning and content survive transmission and reconstruction; they are highly modality-dependent.

- **Image/Video Quality:** Common metrics evaluate image and video quality. Peak signal-to-noise ratio (PSNR) quantifies reconstruction quality by measuring the ratio of maximum signal power to corrupting noise power, assessing pixel-level similarity. Some JSCC frameworks use it as a primary metric for image recovery. In contrast, the structural similarity index measure (SSIM) is a perceptual metric quantifying quality degradation by comparing luminance, contrast, and structure. SSIM can be incorporated into the loss function for training deep JSCC frameworks to directly optimize perceptual quality [52]. Beyond pixel-level and structural metrics, *learned perceptual metrics* are increasingly adopted in SC evaluation. The Learned Perceptual Image Patch Similarity (LPIPS) computes distance in the feature space of a pre-trained deep network, capturing perceptual differences that PSNR and SSIM miss; it is particularly relevant for generative SC systems (Section 6.1) where reconstructions may be perceptually plausible but pixel-misaligned. The Fréchet Inception Distance (FID) measures the distributional similarity between generated and real images, providing a population-level quality assessment. Both metrics connect directly to the rate-distortion-perception trade-off formalized by Blau and Michaeli [81]: optimizing for PSNR forces divergence from perceptual quality (as measured by LPIPS/FID), making the choice of metric a design decision with architectural consequences.
- **Text Semantic Similarity:** Text semantic similarity metrics evaluate meaning preservation after transmission. Traditional metrics like Bilingual Evaluation Understudy (BLEU) assess quality by comparing n-gram overlaps between original and received sentences, but may not fully capture meaning if synonyms are used. A more

advanced approach uses deep learning models, like sentence-BERT, to map sentences into high-dimensional vector embeddings [87]. Semantic similarity is then the cosine similarity between the original and received sentence embeddings. The resulting score (0 to 1) offers a precise measure of meaning preservation and can serve as a key performance indicator or quality constraint.

- **Speech Quality:** Evaluation of speech SC systems depends on the goal: accurate transcription (speech-to-text) or high-fidelity reconstruction (speech-to-speech). For speech-to-text, word error rate (WER) is a standard metric measuring accuracy by counting word substitutions, deletions, and insertions against a ground-truth text [88]. To better capture meaning, semantic similarity scores from sentence embeddings (e.g., BERT) can be used. For speech-to-speech, both objective and subjective metrics are used. Objective metrics include the signal-to-distortion ratio (SDR), the L2 error between the original and recovered signals [89], and Mel cepstral distortion (MCD), a distance between speech spectra [88]. Perceptual metrics approximate human judgment, like the ITU-T standard perceptual evaluation of speech quality (PESQ) [90]. Subjective listening tests yield a mean opinion score (MOS), where human raters score the recovered audio's naturalness [91].

4.2. Metrics for Task Effectiveness and Goal Achievement (Level C)

For task-oriented specializations of SC, metrics at the effectiveness level are central, as they directly quantify the successful completion of a given task. These metrics are inherently goal-oriented and task-specific, with success being defined by the application's ultimate objective. This subsection details the primary metrics that operate at this level, starting with the Value of Information (VoI).

4.2.1. Value of Information (VoI)

VoI is a goal-oriented semantic metric: it quantifies what a piece of information is worth for achieving a specific task. Unlike content-agnostic metrics, VoI is inherently content-aware and task-dependent. It is formally defined as the expected improvement in task performance gained from receiving new information, often expressed as the difference in the expected value of a utility function before and after the information is acquired. Because it directly measures task-relevance, VoI is well suited to optimizing resource allocation and decision-making. Examples of its formulation include:

- **Control Systems:** In a robotic control task, such as a UAV navigating to a waypoint, the VoI of a command can be directly tied to its impact on goal completion. For instance, it could be formulated as the reduction in the predicted distance to the target [92]. A command that corrects a large deviation has a higher VoI than a minor adjustment made while already on course.
- **Classification Tasks:** In a remote visual classification task (e.g., medical diagnosis), the VoI can be linked to model uncertainty. If a low-resolution image results in a low-confidence prediction (high entropy), the VoI of receiving a higher-resolution version is high, as it is expected to significantly reduce classification error. If the model is already highly confident, the VoI of additional data is low.
- **Object Detection:** For an autonomous vehicle, the VoI of detecting an object can be weighted by its class and kinematics. The VoI of correctly identifying a pedestrian crossing the road is far greater than that of a distant, stationary building. This allows the system to prioritize processing and bandwidth for safety-critical information.

As it captures task-relevance, VoI can be combined with other metrics like the Age of Information for intelligent decision-making, such as reordering a command queue at a robot receiver to prioritize high-value, fresh information [92].

4.2.2. Metrics for Task-Specific Success

For many applications, task success is quantified by direct, objective performance indicators. These metrics are highly specific to the task at hand.

- **Classification/Recognition Tasks:** A primary metric for classification and recognition tasks is classification accuracy, the proportion of correctly identified samples. Performance is often optimized by minimizing a loss function, such as cross-entropy. For instance, recent work on multi-task SC uses classification accuracy to evaluate image classification, comparing methods based on coding rate reduction against cross-entropy minimization, especially under label corruption [52].

- **Object Detection Tasks:** Object detection evaluation assesses both classification and localization accuracy. Localization is measured by Intersection over Union (IoU), the overlap between predicted and ground-truth bounding boxes. The primary performance metric is mAP, the average of Average Precision (AP) scores across all object classes; AP summarizes a class's precision-recall curve. mAP is typically calculated at a specific IoU threshold (e.g., mAP@0.5) or averaged over a range of thresholds, and often reported separately by object scale. Optimization involves a multi-task loss combining classification and bounding box regression losses. For example, a UAV object detection study might use MS COCO standards, with mAP at a 0.5 IoU threshold (AP_{50}) as the primary metric, while also considering performance on small (AP_S), medium (AP_M), and large (AP_L) objects to show robustness [93].
- **Reconstruction/Estimation Tasks:** For 3D reconstruction tasks, evaluation focuses on geometric accuracy by measuring positional deviation between original and reconstructed models. For articulated objects like human avatars, assessments calculate the error of individual joints or keypoints, while overall scene distortion is measured using 3D comparison metrics. For instance, a study on AR avatar pose recovery evaluates accuracy at two levels: using mean per joint position error (MPJPE) for skeleton joint errors, and a point-to-point (P2Point) error metric for the overall geometric difference in the final rendered scene [64].

4.2.3. Metrics for Timeliness and Information Freshness

For dynamic, real-time scenarios, the timeliness of information is critical to task effectiveness.

- **Age of Information (AoI):** AoI quantifies information freshness at the destination, defined as the time elapsed since the generation of the most recently received status update. Its main shortcoming is that it is content-agnostic; it penalizes outdated information even if it remains correct.
- **Age of Incorrect Information (AoII):** To address the limitation of AoI, AoII was introduced as a semantics-empowered metric combining age and correctness. AoII is a penalty computed as the product of an information-penalty function (capturing state error) and a non-decreasing time function (increasing the penalty as the mismatch persists) [94]. It effectively measures the time for which the destination has an incorrect understanding of the source's state.

4.2.4. Metrics for User-Centric and System-Wide Performance

Effectiveness can also be measured from a user-centric perspective or at the scale of an entire network.

- **Quality-of-Experience (QoE):** QoE is a user-centric metric quantifying subjective satisfaction. In SC, it is often a composite metric that balances trade-offs between task effectiveness and communication efficiency, reflecting diverse user priorities. The final QoE score is often calculated using functions (e.g., logistic) that map objective metrics to a satisfaction score (0 to 1). For instance, one model for SC networks bases QoE on two components: semantic accuracy (how well the task was performed) and semantic rate (SI transmitted per second, in suts/s) [95]. A weighted sum of satisfaction scores for these components calculates the final QoE, capturing user preferences for either high accuracy or a higher semantic rate.
- **System Throughput in Message (STM):** Traditional throughput in bits is unsuitable for semantic networks as it does not capture the successful delivery of meaning. Instead, network-level metrics are needed to measure total semantic throughput. STM has been proposed to measure the aggregate message rate successfully received by all users in a network [96]. It uses a bit-rate-to-message-rate (B2M) transformation function to convert a link's achievable bit rate into an achievable message rate. This B2M transformation incorporates a semantics-relevant term that accounts for the background knowledge and inference capability of the semantic coding models. By maximizing STM, network resources like user association and bandwidth can be allocated to optimize overall network efficiency from a semantic perspective.

4.3. Metrics for Communication Efficiency and Resource Utilization

A key motivation for SC is improving communication efficiency. Quantifying these resource savings is essential for demonstrating the superiority of semantic approaches over traditional systems.

- **Bandwidth/Data Rate:** Bandwidth and data rate are fundamental efficiency measures, quantified as bits per pixel (bpp) for images, bandwidth reduction percentage, or compression ratio [97]. To measure the transmission of meaning, specialized metrics based on the semantic unit (sut) have been proposed. These include the semantic transmission rate (S-R), which measures effective semantic information transmitted per second (suts/s), and semantic spectral efficiency (S-SE), the rate of successful semantic transmission per unit bandwidth (suts/s/Hz), derived by normalizing S-R by channel bandwidth [87].

- **Latency:** For time-sensitive systems like real-time tasks, end-to-end latency, the total time from source data generation to final task execution, is a critical metric. It comprises processing latency (e.g., semantic encoding/decoding) and transmission latency (influenced by data volume, coding rates, retransmissions). The objective is to minimize this latency while meeting reliability and accuracy requirements. For systems with parallel streams, minimizing the maximum latency among all streams is key to avoiding straggler effects. For example, a generative SC framework might transmit semantic modalities, like a textual prompt and other signals, in parallel [98]. Optimization could involve a reliable retransmission scheme for the critical prompt and adaptive modulation for other signals, with a latency-aware resource allocation scheme assigning power to minimize the maximum delay under semantic quality constraints.

4.4. Metric Conflicts, Trade-offs, and Practitioner Guidance

The metrics defined above are not independent; in practice, optimizing one metric frequently degrades another. Understanding these conflicts is essential for practitioners designing SC/ToSC systems. We identify three categories of metric tension, visualized in Fig. 6, and provide guidance for resolving them.

Level B vs. Level C conflict. Semantic fidelity and task effectiveness can directly conflict: PSNR/SSIM-optimized systems retain details irrelevant to classifiers, while accuracy-optimized systems may produce reconstructions unintelligible to human operators. *Practitioner guidance:* When the system must serve both human inspection and automated tasks, adopt a multi-objective loss with an explicit Pareto trade-off parameter, and report both Level B and Level C metrics. Single-metric optimization is appropriate only when the use case is purely automated. As a concrete illustration, Lyu et al. [52] demonstrate this trade-off in a multi-task image transmission system: sweeping the loss-balance parameter traces a qualitative accuracy–PSNR trade-off curve (their Fig. 10) on which gains in classification accuracy come at the cost of reconstruction PSNR and vice versa, so the two objectives cannot be maximized at a single operating point and are not freely substitutable. Choosing the operating point on this frontier is a design decision rather than a property the system can optimize away.

Task effectiveness vs. efficiency (timeliness) conflict. Maximizing VoI or accuracy may require additional data or complex inference, increasing latency and degrading AoI. For example, in UAV command-and-control, a stale but high-value command is counterproductive: value-of-information and age-of-information objectives must be traded off so that timely, high-value commands take precedence [92]. *Practitioner guidance:* Define a composite utility function that explicitly penalizes both task error and staleness, such as AoII-weighted accuracy. Prioritize AoI minimization for safety-critical decisions and accuracy maximization for non-time-critical analytics within the same system using a priority queue based on VoI.

Efficiency vs. fidelity (robustness) conflict. Aggressive semantic compression reduces bandwidth but also reduces the redundancy that protects against channel errors and KB mismatch. A system operating at maximum semantic spectral efficiency (S-SE) has no margin for error recovery if the receiver's KB diverges from the sender's. *Practitioner guidance:* Design compression rates with explicit margin for the expected KB mismatch level and channel variability. In dynamic environments, adaptive compression that increases redundancy when confidence is low (measured by receiver feedback) is preferable to fixed-rate schemes.

Limitations of current metrics. Two gaps remain in current evaluation practice. First, VoI is theoretically compelling but rarely computable in closed form for realistic scenarios; most implementations use proxy metrics (e.g., prediction confidence) that may not faithfully represent true task value. Second, no existing metric captures *cross-task generalizability*, the degree to which a semantic representation supports not just the intended task but also related tasks that may arise during deployment. Developing such a metric is an open research challenge that directly affects the practical viability of ToSC systems in multi-task environments.

The need for standardized evaluation. Underlying all of these tensions is a methodological obstacle: there is no standardized way to evaluate SC/ToSC systems, so reported gains are frequently incomparable across papers. As our cross-domain analysis shows (Pattern 4, Section 5), works differ in datasets, channel models, compression baselines, and even the definition of "task success," which makes it impossible to determine whether a new system genuinely

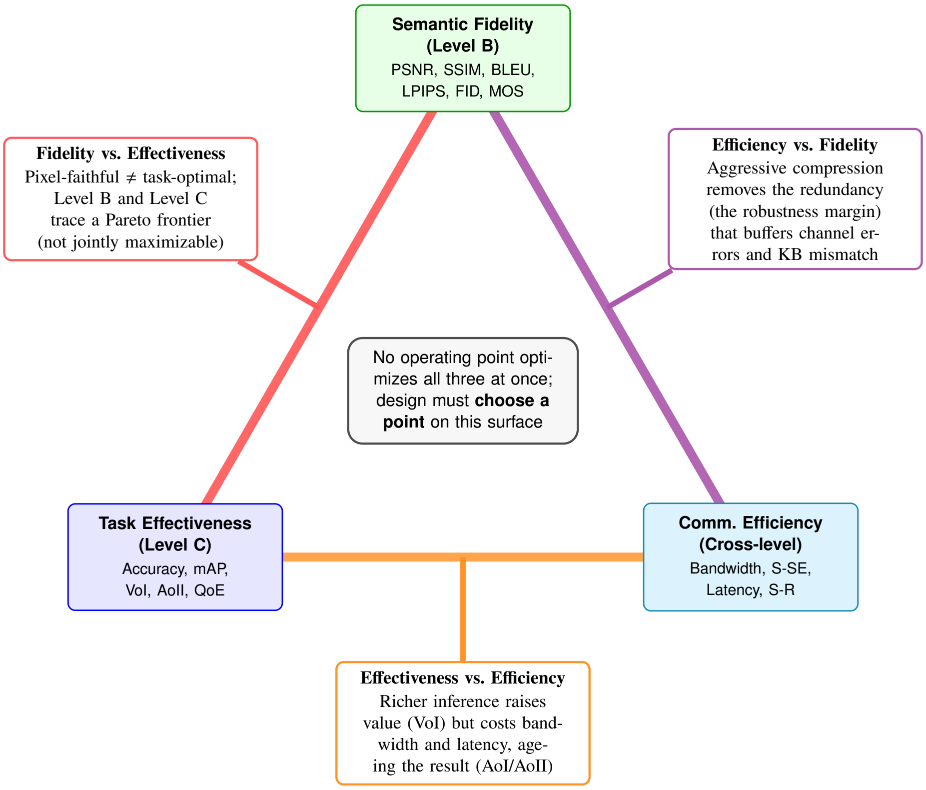

Figure 6: Metric conflict and trade-off map for SC/ToSC evaluation. The three vertices are the primary metric families: semantic fidelity (Level B), task effectiveness (Level C), and communication efficiency (cross-level). Each edge denotes a pairwise conflict with its governing mechanism: fidelity vs. effectiveness (the Level B and Level C objectives trace a Pareto frontier and cannot be jointly maximized), effectiveness vs. efficiency (richer inference raises value-of-information but costs bandwidth and increases age-of-information), and efficiency vs. fidelity (aggressive compression removes the redundancy, i.e. the robustness margin, that buffers channel errors and knowledge-base mismatch). As the central node states, no single operating point optimizes all three simultaneously, so system design reduces to selecting a justified point on this surface.

advances the state of the art or merely benefits from a weaker baseline or an easier channel. Addressing this requires three coordinated efforts. First, a *unified benchmark* (public, multi-modal datasets carrying both reconstruction ground truth and task annotations) so that Level B fidelity and Level C effectiveness can be measured on the same footing. Second, a *standardized test protocol* that fixes the channel models, the compression baselines (including strong neural codecs, not only JPEG/BPG), the SNR operating points, and the reporting requirements, so that results are reproducible and directly comparable. Third, an *open evaluation platform*, namely a public leaderboard with reference implementations and held-out test channels, that additionally reports *deployment-oriented* measures (inference latency, model size, energy per inference, and robustness under KB mismatch) rather than task accuracy alone, so that practical deployability, not benchmark accuracy in isolation, becomes a first-class evaluation criterion. We develop this as a prioritized engineering direction in Section 7 (Priority 2).

5. Use Cases in ToSC

The principles of ToSC address key challenges in high-impact technological domains. In areas from mobile robotics to industrial automation, the core challenges are similar: maintaining reliable, low-latency performance in complex, resource-constrained environments. ToSC reduces communication overhead and energy consumption not through simple compression, but by extracting and transmitting only task-essential information. Table 5 provides a consolidated comparison of representative works, and the subsections below examine how ToSC enhances system performance in each domain, with explicit analysis of task-specific requirements, semantic information extraction, and performance metrics. Each subsection concludes with a critical assessment that identifies domain-specific limitations, negative results, and unresolved challenges.

5.1. Autonomous Vehicular and Vehicular Communication Networks

Autonomous transportation is a widely studied use case for ToSC. An autonomous vehicle must process sensor data in real time for navigation, obstacle avoidance, and traffic compliance, and transmitting complete raw sensor streams introduces latency and bandwidth overhead that these functions cannot tolerate. What actually feeds a driving decision is far sparser: recognized traffic signs, detected obstacles, and planned trajectories. ToSC extracts and transmits those elements and discards the rest; success is judged at the driving level, by safety (accident rates), navigation accuracy, and real-time response under varying traffic conditions.

Recent work demonstrates ToSC's potential in driving scenarios, though important caveats apply. Eldeeb et al. [53] proposed a multi-task SC framework for autonomous vehicles that extracts semantic features from traffic signs for both reconstruction and classification tasks, achieving up to an 89% reduction in transmitted bits while maintaining high task performance. However, serving several driving subtasks from one encoder requires keeping their features separable: the unified multi-task system of Zhang et al. [67] employs domain adaptation precisely to project each task's features into its own domain and transmit only the task-specific ones, a design burden that grows when additional driving subtasks (lane detection, depth estimation) are added. To handle multi-user and multimodal data, Feng et al. [99] developed a system for edge-enabled autonomous driving that fuses semantic features from multiple vehicles and data types at an edge server, reducing image processing runtime by over 95% compared to traditional JPEG methods. Lv et al. [100] focused on efficient communication by developing an importance-aware SC system for vehicular image segmentation. Their VIS-SemCom system prioritizes critical objects like vehicles and pedestrians, cutting transmitted data by about 70% at no cost to segmentation quality. However, all three systems define "task-relevant" features at training time, a definition that can be brittle when unfamiliar obstacle types or unexpected road conditions fall outside the training distribution; the adaptable-compression framework of Liu et al. [101] addresses a complementary axis, jointly adapting the semantic compression ratio and wireless resources to multi-user delay constraints.

Critical assessment. Despite encouraging results, the cited studies share a common limitation: evaluation under controlled or simulated channel conditions. No work has demonstrated ToSC reliability under real-world vehicular channel dynamics (fast fading, Doppler spread, handover). All systems also assume a fixed task taxonomy (predefined sign classes, known obstacle types); performance under open-set conditions (where previously unseen objects appear that were not in the training distribution) remains untested. The safety-critical nature of autonomous driving demands formal guarantees on worst-case performance that no existing ToSC system provides.

5.2. Satellite and UAV-Aided Communication Networks

Satellite and UAV links are where bandwidth scarcity is most clearly a physical limit: a satellite downlink pass delivers a fraction of what the onboard sensors capture, and a UAV's energy budget caps both its computation and its transmissions. The tasks these platforms serve (Earth observation, remote sensing, surveillance missions, real-time control of aerial platforms) therefore force a choice about what crosses the link: geospatial features, detected anomalies, mission-critical status updates, or compressed observations that keep only what analysis needs. Evaluation centers on mission completion rates, data quality retained under the bandwidth budget, and latency for time-critical operations. The surveyed systems split by platform: satellite work concentrates on downlink compression under extreme bandwidth constraints, UAV work on uplink efficiency with energy-constrained onboard processing. Both rely on task-driven feature selection, but differ in their handling of KB synchronization: satellite systems assume ground-station KBs are static, whereas UAV swarms require dynamic KB updates across mobile agents.

Hassan et al. [102] explore SC for resource-constrained 6G LEO satellite communication in Earth observation missions. They propose a framework that uses joint semantic-channel encoding and optimization algorithms to minimize latency in imagery downlinks while meeting quality-of-service constraints. Results show their scheme effectively reduces latency and compression ratios compared to baselines, while consistently preserving required data quality. For FSO-based satellite communication, Chen et al. [103] propose an FSO-SC scheme for remote sensing image transmission. Their method uses a VQ-VAE to extract key semantic features. Experiments demonstrated a power gain of over 3 dB and a 60% reduction in communication overhead compared to traditional systems, all while

Table 5: Comparison of representative ToSC works across ten application domains. Reported gains are each work's own comparison, predominantly measured against traditional codecs (e.g., JPEG/BPG) or raw/full-state transmission rather than a state-of-the-art learned codec; see the weak-baseline caveat (Pattern 2, Section 5).

Domain	Reference	Core Technique	Key Result
Vehicular	Eldeeb et al. [53]	Multi-task SC for traffic sign recognition/classification	89% transmitted bit reduction
	Zhang et al. [67]	Unified multi-task encoder backbone	Comparable accuracy to single-task models with much less transmission overhead
	Feng et al. [99]	Edge-enabled multimodal semantic fusion	95%+ image processing runtime reduction
	Lv et al. [100]	Importance-aware SC (VIS-SemCom)	~70% data reduction, high segmentation quality
	Liu et al. [101]	Adaptable-ratio SC with resource allocation	Jointly adapts compression ratio and wireless resources to multi-user delay constraints
Sat/UAV	Hassan et al. [102]	Joint semantic-channel encoding for LEO satellites	Reduced latency while preserving data quality
	Chen et al. [103]	FSO-SC with VQ-VAE for remote sensing images	3 dB power gain; 60% overhead reduction
	Xu et al. [104]	Task-oriented transmission with UoI metric	Minimized UoI and power consumption
	Song et al. [105]	Cognitive SC with knowledge graph for UAV detection	Only 0.6% accuracy drop under noisy channels
	Xu et al. [106]	Hierarchical federated learning for UAV swarm SC	Higher PSNR and lower MSE than baseline
	Xu et al. [107]	TOSA framework with DRL for C&C packets	Near-constant transmit count vs. linear growth; matched effectiveness
Surveillance	Shah et al. [108]	PeR-ViS: semantic description-based person retrieval	High IoU and detection rate, surpassed prior best
	Shao et al. [109]	TOCOM-TEM: task-oriented edge video analytics	62–85% comm. cost reduction at matched accuracy
	Nandakishore et al. [110]	Background-as-shared-knowledge, dynamic ROI only	86%+ bit savings with improved PSNR
	Huang et al. [111]	SPE: privacy-preserving SC via semantic scene graphs	Parameter-only exchange (~50 MB vs. up to 50 GB raw)
IoT/IIoT	Yang et al. [112]	Task-relevant semantic extraction for AI-task classification	>40% accuracy gain; much lower bandwidth
	Xie et al. [113]	L-DeepSC: pruned/quantized SC for IoT	Model 90% smaller with minimal perf. loss
	Hu et al. [114]	Robust SC under semantic noise	Masked VQ-VAE codebook robust to adversarial semantic noise
	Chen et al. [115]	Edge-assisted optimal model selection per device	Near-optimal solution in simulations
	Moqurrab et al. [116]	Instant_Anonymity: perturbation-based privacy	Lower query error vs. k-anonymity
	Willner et al. [117]	oneM2M + Semantic Web for Industry 4.0	Integrates existing formal information models
	Zhao et al. [118]	Distributed RIS with probabilistic SC	Up to ~2× semantic-aware rate vs. centralized RIS
	Zhang et al. [119]	ISC with semantic flags for tool wear monitoring	12.9% delay reduction
Healthcare	Yang et al. [120]	ISCSC: joint beamforming + semantic extraction	70% transmission rate improvement
	Bura et al. [121]	DeepSemECG for robust biomedical signal SC	Outperformed baseline on CDI under noise
	Yang et al. [122]	Superimposed SC for real-time ECG monitoring	98.8% accuracy at 50× compression
Robotics	Li et al. [123]	ToSC + GenAI agent for multi-robot anomaly detection	Data reduced to 2.5% of raw images
	Wu et al. [124]	H2R Bridge: few-shot human intent via a fine-tuned vision-language model	92%+ accuracy with 5 examples per action
	Suresh et al. [125]	XR-integrated SC for human–multi-robot exploration	Underlying SC operates at a rate-normalized SNR up to 20 dB lower than classical
Digital Twin	Chen et al. [126]	Goal-oriented SC with feature & temporal selection	59.5–80% comm. load reduction (sim.)
	Tang et al. [127]	DRL-based compact SC for UAV-assisted DT sync	8.23% sync delay; 57% energy savings
	Li et al. [128]	Joint optimization of sync, power, and resources	13.2% sync latency reduction
	Kreuzer et al. [129]	Review of AI in digital twins (incl. predictive maintenance)	Surveys proactive-maintenance techniques
Metaverse/XR	Huang et al. [130]	ISCom: interest-aware SC for point cloud streaming	10–22 FPS improvement on mobile
	Yang et al. [131]	Streamlined SC + GenAI for multi-modal XR data	Efficient heterogeneous data transmission
	Chen et al. [74]	Cross-modal SC with GNNs and cross-modal attention	High cross-modal semantic similarity across SNRs
	Wang et al. [64]	Goal-oriented SC for avatar synchronization	95.6% latency reduction
	Liu et al. [132]	Vision-based SC encoding skeleton coordinates	51 bytes vs. 8.243 MB per frame
	Zhang et al. [133]	LLM-integrated SC for context-aware 6G XR	Goal-directed bandwidth allocation
	Li et al. [134]	Federated cloud-edge-end SC for privacy-preserving XR	96% delay reduction; 44% quality gain
Text SC	Xie et al. [37]	DeepSC: Transformer-based end-to-end text SC	Superior similarity vs. Huffman+LDPC
	Xie et al. [113]	L-DeepSC: lightweight SC via pruning/quantization	12.3 MB to 1.28 MB model size
	Xie et al. [12]	MU-DeepSC: multi-user SC for text + image (VQA)	Joint multi-modal multi-user optimization
	Zhou et al. [41]	Knowledge graph-based SC with S-P-O triples	Structured semantic transmission + reasoning
	Liu et al. [135]	Triplet-based explainable SC with filtering	Reduced data, maintained interpretability
	Shi et al. [136]	RIS-assisted JSCC for text	Improved BLEU at low SNR
Speech SC	Weng and Qin [89]	DeepSC-S: attention-based end-to-end speech SC	Higher SDR and PESQ than separate coding at low SNR
	Han et al. [88]	SC preserving semantic content + speaker identity	Competitive PESQ/MOS, reduced data

maintaining image quality. To improve efficiency in Satellite-Integrated Internet, Xu et al. [104] introduce a task-oriented transmission (TUT) framework. Their "utility loss of information (UoI)" metric, which captures freshness and task status, guides the semantic coding. Simulations showed their approach effectively minimizes long-term UoI and power consumption, outperforming several benchmarks.

On the UAV side, where agile, low-latency control dominates the design requirements, Song et al. [93] proposed a cognitive SC system for UAV-based object detection that uses a knowledge graph (KG) to improve performance with limited onboard resources. A semantic extractor on the UAV sends compressed data, and the KG at the ground server enhances semantic interpretation; across low-SNR AWGN and Rayleigh channels their evaluation reports consistent mean-average-precision gains over an attention-based deep JSCC baseline. An earlier version of this work [105] reported superior resilience under noisy channel conditions, with a minimal 0.6% accuracy drop compared to over 3% for baseline systems. To train semantic models in a UAV swarm, Xu et al. [106] designed an SC architecture using hierarchical federated learning (HFL). This design improves reliability by organizing UAVs into clusters for decentralized training. Experiments showed the HFL-powered model significantly outperformed the baseline under dynamic conditions, achieving higher PSNR and lower mean squared error (MSE). Xu et al. [107] developed a task-oriented semantics-aware (TOSA) framework to improve efficiency when sending command and control (C&C) signals to UAVs. A deep reinforcement learning (DRL) agent learns to drop redundant or less timely packets based on their value. The learned policy cuts resource consumption sharply: it transmits only the most valuable command-and-control signals, keeping the number of transmissions nearly constant as packet repetition grows while achieving nearly the same effective transmission rate as bit-oriented transmission of every packet.

Critical assessment. Satellite and UAV ToSC systems face a unique challenge: the extreme asymmetry between uplink and downlink capabilities makes KB synchronization difficult. Most frameworks assume the ground station's KB is perfectly aligned with the onboard encoder, but orbital dynamics, firmware update cycles, and mission-specific reconfiguration introduce persistent KB drift. The energy constraints of small satellites and UAVs impose hard limits on encoder complexity, yet most proposed systems rely on computationally intensive DNNs. How the reported laboratory performance translates to space-qualified hardware is a question the surveyed work has yet to address.

5.3. Video Surveillance

Large-scale video surveillance generates prohibitive volumes of raw video, yet its downstream tasks (person identification, activity recognition, anomaly detection, threat assessment) depend not on pixel-level fidelity but on sparse cues: human pose, facial characteristics, motion patterns, behavioral descriptors. ToSC transmits the cues instead of the pixels, cutting data volume without sacrificing detection accuracy, identification precision, or response time.

For person retrieval, Shah et al. proposed PeR-ViS, a system that uses a semantic description (e.g., clothing, gender) as a query instead of an image [108]. Their deep learning approach acts as a cascade filter to identify the target individual, achieving a high average IoU and detection rate that surpassed the best prior result. Shao et al. developed TOCOM-TEM, a task-oriented framework for low-latency edge video analytics [109]. By removing spatial and temporal redundancy, the framework cut the communication cost from roughly 130 KB to 20–50 KB per frame (a 62–85% reduction) at matched accuracy (~85% multiple-object-detection accuracy) for a multi-camera pedestrian occupancy prediction task. In static surveillance environments, Nandakishore et al. introduced a technique that uses the unchanging background as shared knowledge [110]. The system transmits only the dynamic Region of Interest (e.g., a person). This method achieved average bit savings of over 86% and simultaneously improved PSNR compared to a baseline. To address privacy, Huang et al. proposed SPE, a decentralized, semantic privacy-preserving framework for edge nodes [111]. The system generates a semantic scene graph instead of transmitting raw video and trains models by only exchanging parameters. Communication overhead falls accordingly: nodes exchange only model parameters, about 50 MB per node versus local datasets of up to 50 GB, roughly 0.1% of the volume that sharing raw data would require.

Critical assessment. Video surveillance ToSC faces a tension between compression and accountability: discarding "task-irrelevant" visual information at the encoder means the original scene cannot be reconstructed for forensic review or legal evidence. Systems like PeR-ViS that operate on semantic descriptions rather than images inherently lose the ability to verify false positives after the fact. The privacy-preserving approaches (e.g., scene graph transmission) trade one risk for another: while raw video is not transmitted, the semantic graph itself may leak sensitive information about activities and social interactions. The ethical and regulatory implications of semantic-level surveillance remain largely unexplored.

5.4. IoT and Industrial IoT Environments

IoT deployments face a different constraint: enormous device counts with tiny per-device budgets. An IoT network monitors environments, controls devices, aggregates data, and automates decisions, and for each of these the payload that matters is small: an abstracted sensor reading, a status indicator, an anomaly alert, a control command. ToSC strips the raw-data redundancy around those payloads, and is measured on system responsiveness, energy efficiency, network utilization, and task completion accuracy across the deployment. We consider general IoT first and the specialized Industrial IoT (IIoT) setting next.

To address challenges in the AIoT, Yang et al. [112] proposed SC-AIT, a semantic communication architecture that transmits only the task-relevant semantic information and discards content irrelevant to the target task. Implemented for image classification, it achieved much lower bandwidth requirements and over 40% classification-accuracy gains when compared to conventional bit-level (technical-level) communication. Xie et al. [113] developed L-DeepSC, a lite distributed semantic communication system for resource-constrained IoT devices. A model trained in the cloud is compressed via pruning and quantization for execution on devices, reducing the model size by nearly 90% (from 12.3 MB to 1.28 MB) with minimal impact on performance. However, Hu et al. [114] show that conventional semantic codebooks are vulnerable to *semantic noise* (adversarial, sample-dependent and sample-independent perturbations that distort the intended meaning) and propose a masked VQ-VAE codebook to restore robustness, a concern compounded in IoT deployments where heterogeneous devices and time-varying inputs stress the shared codebook. To manage the complexity-accuracy trade-off in edge-assisted SC, Chen et al. [115] proposed a framework where an edge server selects the optimal semantic model for each IoT device's task, achieving a near-optimal solution in simulations. A common limitation across these IoT systems is that they assume stationary data distributions; in practice, sensor drift and firmware updates continuously alter the statistical properties of the data, yet none of the proposed frameworks include mechanisms for online model adaptation.

Smart factories intensify those demands: decisions must be real-time, devices from different vendors must interoperate, and process data is commercially sensitive. IIoT tasks (predictive maintenance, quality control, production optimization, safety monitoring) consume machine health indicators, production quality metrics, and operational status updates, and the payoff is realized in production: reduced downtime, improved quality, lower cost.

For data privacy in 5G-enabled IIoT, Moqurrab et al. proposed Instant_Anonymity, a lightweight semantic privacy framework [116]. The perturbation-based technique is suitable for resource-constrained devices and demonstrated superior data utility with much lower query error compared to the traditional k-anonymity model. Focusing on interoperability, Willner et al. argued for semantic-level communication in smart factories [117]. They proposed using the oneM2M standard, which supports Semantic Web technologies, to implement the Industrie 4.0 Asset Administration Shell concept, enabling the reuse of hundreds of existing formal information models and their linkage to thousands of data sources. To enhance spectral efficiency in IIoT, Zhao et al. investigated a framework combining distributed RIS with probabilistic semantic communication (PSC) [118]. By jointly optimizing system parameters, their distributed RIS schemes achieved up to roughly twice the sum semantic-aware transmission rate of a centralized RIS deployment with the same total number of elements. For tool wear monitoring, Zhang et al. [119] developed an Industrial Semantic Communications (ISC) framework. The system classifies the machine's state and transmits only a semantic flag during irrelevant "idling" periods. This intelligent filtering reduced end-to-end transmission delay by 12.9% compared to a standard scheme.

Critical assessment. IoT/IIoT environments present perhaps the hardest deployment scenario for ToSC: extreme device heterogeneity, limited compute on edge nodes, and the need for long-term (multi-year) model stability. Most cited systems are evaluated on homogeneous device populations with stationary data distributions. In practice, sensor drift, firmware updates, and evolving production processes continuously alter the data distribution, requiring mechanisms for online model adaptation that most current ToSC frameworks do not address, though few-shot adaptive schemes such as that of Han et al. [137] begin to provide such online adaptation, albeit for shifting channel conditions rather than input-data drift. The interoperability challenge is also underestimated: while semantic standards like oneM2M exist, achieving semantic alignment across thousands of heterogeneous devices from different manufacturers remains an open engineering problem.

5.5. *Medical and Healthcare Systems*

Healthcare shows both the promise and the limits of ToSC. A clinician monitoring a patient in real time, reading a scan, planning treatment, or responding to an emergency works from diagnostic features (waveform morphology, biomarker thresholds, abnormality indicators), not from full-fidelity signal reconstruction, so delivering only those features cuts transmission overhead. The difficulty is that in medicine the boundary between "task-relevant" and "task-irrelevant" is less clear-cut than elsewhere: subtle signal features may carry diagnostic significance that is not apparent from the training data.

Yang et al. [120] introduce an Integrated Sensing, Computing, and Semantic Communication (ISCSC) framework for e-Healthcare. By jointly optimizing beamforming and semantic extraction, their approach aims to maximize the Semantic Secrecy Rate (SSR) for privacy and minimize sensing error, improving the transmission rate by up to 70% compared to traditional benchmarks. Bura et al. [121] developed DeepSemComm, a framework for robustly transmitting biomedical signals whose DeepSemECG model preserves the diagnostic integrity of ECG signals in noisy environments, achieving lower Clinical Distortion Index (CDI) than a CNN autoencoder baseline under additive noise. For IoT-based healthcare, Yang et al. [122] propose a framework for real-time ECG monitoring using superimposed semantic communication, achieving a classification accuracy of 98.8% with a compression rate of 50x. However, healthcare spans both regimes identified in Section 3.6. Open-ended diagnostic reading, where a clinician inspects the entire image for any abnormality, approaches the small-gap convergence case ($H(X) - I(X;T)$ small), so ToSC offers little advantage over high-fidelity SC. Targeted detection of a subtle, low-variance abnormality is instead the low-variance, high-value case: the diagnostic target sits within a large task-irrelevant background, so a task-aware encoder *can* outperform reconstruction, but only if it is given a relevance signal identifying the target. Absent such a signal, aggressive task-oriented compression risks discarding clinically relevant information, particularly for rare pathologies absent from the training distribution, where the semantic encoder has no basis for judging relevance.

Critical assessment. The two-regime split above has direct deployment consequences. The reported 50x compression ratio for ECG monitoring, while impressive, has not been validated against the full spectrum of cardiac pathologies; rare arrhythmias that were absent from the training set may be filtered out as "task-irrelevant." Regulatory approval for diagnostic systems that operate on lossy semantic representations (rather than full-fidelity data) remains an unresolved barrier. The security frameworks also focus on data confidentiality but do not address the risk of adversarial semantic attacks that could alter diagnostic outcomes.

5.6. *Robot Collaboration and Human-Robot Interaction*

Multi-robot collaboration and human-robot interaction demand communication that scales with team size. A teammate needs the semantics of the shared goal, not another robot's raw sensor feed: recognized objects, spatial relationships, motion parameters, gesture interpretations. ToSC transmits exactly that, which keeps coordination within the bandwidth budget.

Li et al. [123] introduce a framework combining ToSC with a Generative AI (GenAI) agent to enhance multi-robot systems. In an anomaly detection task, robots use a semantic encoder before transmitting images. This reduced data transmission volumes to as little as 2.5% of that required for raw images without giving up classification accuracy. To address data scarcity in human-robot collaboration, Wu et al. [124] present the H2R Bridge, a framework for understanding human intent from few examples. Their system uses a fine-tuned vision-language model to interpret human actions and translate them into robot commands. With only five training examples per action, it reached over 92% accuracy. For exploration in hazardous environments, Suresh et al. [125] present a framework integrating human operators with a multi-agent robotic team using extended reality and SC. To avoid overwhelming operators, the system transmits only the essential meaning of exploration data. The chosen SC system was previously shown to operate effectively at a rate-normalized SNR up to 20 dB lower than classical communication, ensuring timely decision-making.

Critical assessment. Multi-robot ToSC systems are evaluated primarily on static task definitions with cooperative agents. Real-world multi-robot deployments face adversarial conditions (competing objectives, communication jamming, sensor spoofing) that are not addressed. The H2R Bridge's few-shot intent recognition, while encouraging, relies on a pre-trained vision-language model whose failure modes under domain shift (e.g., from laboratory to industrial environments) are not characterized. The reliance on shared KBs among robots creates a single point of failure: if the KB becomes corrupted or desynchronized, the entire multi-agent system may fail simultaneously rather than degrading gracefully.

5.7. *Digital Twin Synchronization and Predictive Maintenance*

Digital Twin (DT) systems require real-time synchronization between physical assets and their virtual counterparts for predictive maintenance and performance optimization [138]. ToSC addresses the bandwidth saturation inherent in continuous full-state transmission by extracting only task-relevant semantics: state change indicators, degradation signals, and anomaly features. For robotic arm reconstruction in DT environments, Chen et al. [138] proposed a goal-oriented framework that pairs a feature-selection algorithm with deep reinforcement learning-based temporal selection, transmitting only the most informative features and time steps rather than periodic full-state updates; the extended version [126] reduces communication load by at least 59.5% (strict reconstruction-error constraint) and 80% (relaxed constraint) in simulation while maintaining reconstruction accuracy.

Tang et al. [127] proposed a framework using compact SC systems to minimize latency in UAV-assisted DT synchronization. By transmitting only essential information, their DRL-based optimization algorithm reduced DT sync delay by 8.23%, data dropping rates by 15.31%, and energy consumption by up to 57.14% compared to benchmarks. For mobile edge networks, advanced semantic communication frameworks address the mobility of user devices and dynamic channel conditions, achieving up to 13.2% reduction in synchronization latency through joint optimization of synchronization strategy, transmission power, and computational resource allocation [128]. Okegbile et al. [139] instead target the security of this synchronization, integrating split federated learning with task-oriented secure semantic communication and a token-based semantic-defense mechanism that distinguishes adversarial from authentic semantic data during physical-to-virtual updates.

For predictive maintenance, AI-driven digital-twin systems already base proactive decisions on degradation signatures and failure precursors (vibration anomalies, temperature deviations, wear metrics) [129, 140, 141]; a ToSC formulation would transmit only these precursors rather than complete sensor streams, keeping communication overhead minimal.

Critical assessment. DT synchronization presents a unique temporal challenge: the "ground truth" physical state evolves continuously, but semantic updates are discrete and lossy. Current systems report average synchronization improvements but do not analyze worst-case desynchronization events, precisely the scenarios where predictive maintenance decisions are most consequential. The feature- and temporal-selection approach, while effective for the repetitive motion of robotic-arm reconstruction, remains untested on irregular or previously unseen motion patterns. A deeper gap remains: no formal definition of "semantic synchronization fidelity" exists that captures not just state accuracy but the impact of synchronization errors on downstream maintenance decisions.

5.8. *Metaverse and Extended Reality Applications*

Metaverse and Extended Reality (XR) applications stream visual, audio, haptic, and spatial content in real time, yet only a small fraction of it matters to the user at any instant. ToSC exploits exactly that, extracting user attention, interaction intentions, avatar motion, and environmental context instead of complete sensory streams [64].

The works in this domain span three architectural strategies: (1) region-of-interest selection that compresses spatial data by transmitting only user-attended regions (ISCom, skeleton-based systems); (2) cross-modal fusion that jointly encodes multiple modalities through shared semantic representations (graph neural network and Transformer-based approaches); and (3) foundation model-assisted systems that exploit LLM reasoning for context-aware bandwidth allocation. These strategies trade off compression aggressiveness against reconstruction controllability, with the most aggressive approaches (skeleton-only transmission at 51 bytes) achieving orders-of-magnitude compression but sacrificing the ability to reconstruct visual detail for verification.

For XR device optimization, the interest-aware ISCom framework improves point cloud video streaming, raising the minimum rendering frame rate by 10 FPS and up to 22 FPS on mobile devices through region-of-interest selection and lightweight encoder-decoder architectures tailored to resource-constrained XR devices [130]. Yang et al. integrate semantic awareness with generative AI to achieve streamlined transmission, using multi-modal fusion and superposition schemes for heterogeneous data [131].

Cross-modal semantic communication suits Metaverse applications because it integrates text, audio, image, and haptic modalities through graph neural networks and Transformer-based encoders. These systems compress semantic features through cross-modal attention mechanisms while using generative AI for multimodal data reconstruction in virtual scenarios [74]. Total transmitted data decreases, and cross-modal semantic similarity stays high under the authors' own similarity measure across the reported channel conditions.

For avatar synchronization, goal-oriented frameworks achieve up to 95.6% latency reduction [64], while skeleton-based systems compress avatar data to just 51 bytes [132]. LLM-integrated approaches enable context-aware bandwidth allocation for immersive experiences [133], and federated cloud-edge-end architectures reduce transmission delay by 96% while addressing privacy concerns through distributed semantic processing and 3D scene reconstruction [134]. The skeleton-based and LLM-integrated systems instantiate strategies (1) and (3) above, while the avatar-synchronization and federated cloud-edge-end frameworks apply the same task-relevance lever to latency and privacy objectives.

Critical assessment. Metaverse/XR applications demand extremely low latency (e.g., sub-20 ms transmission budgets for augmented reality [64]) while simultaneously requiring high visual fidelity for user comfort. Current ToSC systems report impressive compression ratios (e.g., 51 bytes for skeleton data) but do not evaluate perceptual quality from the user's perspective; metrics like MPJPE may be acceptable numerically while producing visually jarring artifacts that break immersion. The reliance on pre-shared base knowledge (e.g., an initial avatar mesh) creates a cold-start problem for new users or environments. The LLM-assisted approaches introduce inference latency that may conflict with the stringent real-time requirements of XR applications.

Text and speech SC, the two modality-level domains, are the last of the ten; their techniques underpin many of the application-oriented systems above.

5.9. Text-Based Semantic Communication

Text-based SC, brought to prominence by the foundational DeepSC system, targets sentence-level meaning preservation across noisy channels for applications including machine translation, dialogue, and information retrieval. Performance is evaluated through Level B metrics (BLEU, sentence similarity, WER), though the gap between n-gram overlap and true meaning preservation motivates embedding-based measures (Section 4).

Building on earlier deep-learning JSCC for text [142], Xie et al. [37] introduced DeepSC, the first Transformer-based end-to-end SC system for text, jointly optimizing semantic encoding and channel coding. DeepSC demonstrated that transmitting learned semantic features rather than source-coded bits achieves superior sentence similarity, particularly at low SNR, compared to conventional source-channel separation schemes such as Huffman coding combined with low-density parity-check (LDPC) coding. Subsequent work extended this design in several directions. Xie et al. [113] developed L-DeepSC, a lightweight version using pruning and quantization to reduce the model from 12.3 MB to 1.28 MB for resource-constrained IoT deployment. For multi-user scenarios, MU-DeepSC [12] enables simultaneous text and image transmission for tasks such as VQA, with jointly optimized transceivers. Zhou et al. [41] proposed representing text semantics via knowledge graphs, specifically subject-predicate-object triples, enabling explicit, structured semantic transmission that supports reasoning at the receiver. The triplet-based explainable SC scheme of Liu et al. [135] further developed this direction by filtering redundant triplets to reduce data volume while maintaining interpretability.

A distinct architectural branch applies JSCC to text by mapping sentences directly to channel symbols without intermediate bit representations. The RIS-assisted text SC system of Shi et al. [136] demonstrated that jointly optimizing semantic encoding with RIS phase shifts improves BLEU scores in low-SNR conditions, illustrating how physical-layer enhancements complement semantic-level design.

Critical assessment. Text SC benefits from the maturity of NLP models, but several limitations persist. Most systems are evaluated on short, well-formed sentences from curated datasets; performance on noisy, informal, or domain-specific text (medical reports, legal documents) is largely uncharacterized. The reliance on pre-trained language model weights as implicit (parametric) knowledge creates the same KB alignment weakness identified across other domains: if the transmitter and receiver use different model versions or fine-tuning, semantic decoding degrades. Evaluation is dominated by BLEU and sentence similarity, neither of which captures pragmatic correctness, i.e., whether the received text leads to the correct downstream action. No benchmark yet evaluates text SC under realistic multi-turn dialogue conditions with evolving context.

5.10. Speech and Audio Semantic Communication

Speech SC targets both speech-to-text (transcription) and speech-to-speech (reconstruction) scenarios, transmitting phonemic content, prosodic features, and speaker identity. Performance spans Level B (WER, PESQ, MOS) and Level C (downstream task accuracy for classifiers or command interpreters).

Weng and Qin [89] introduced DeepSC-S, among the first deep learning-based SC systems for speech, using an attention-based encoder-decoder trained end-to-end to extract and transmit speech semantics. DeepSC-S demonstrated that semantic-level transmission achieves higher signal-to-distortion ratio and perceptual speech quality than conventional separate source-channel coding, especially at low SNR, by learning squeeze-and-excitation attention weights that emphasize the essential speech information. Han et al. [88] further improved speech SC efficiency by proposing a system that preserves both semantic content and speaker characteristics, achieving competitive PESQ and MOS scores while reducing transmitted data volume. The system exploits the observation that much of the speech signal is acoustically redundant for meaning extraction (pauses, filler sounds, and predictable phonetic transitions), enabling aggressive compression without semantic loss.

A recent direction integrates neural audio codecs (e.g., EnCodec [143], which uses residual vector quantization) into the SC pipeline. These codecs learn discrete token representations of audio that are suited to standard digital transmission while preserving perceptual quality, bridging the analog JSCC and digital modular paradigms identified in Section 3.3.

Critical assessment. Speech SC shares the knowledge dependency challenge: encoder-decoder pairs must be trained on matched data distributions, and performance degrades when the speaker population, language, or acoustic environment at deployment diverges from the training set. We found no work addressing multilingual or code-switching scenarios, though both are common in real-world communication. Evaluation under realistic acoustic conditions (background noise, reverberation, multi-speaker overlap) is sparse. Most critically, speech SC for task-oriented applications (voice commands for robotics, diagnostic interviews in telemedicine) has received minimal attention; the majority of work focuses on reconstruction fidelity rather than downstream task performance, leaving the SC/ToSC distinction for speech largely unexplored.

5.11. Cross-Domain Synthesis: Recurring Patterns and Lessons

The critical assessments above reveal several cross-cutting patterns that recur across individual application domains:

Pattern 1: knowledge dependency is the dominant bottleneck. Nearly every domain depends on aligned knowledge (an explicit shared KB where one is maintained, or, for the parametric text- and speech-SC systems, the implicit knowledge embedded in model weights), yet none has solved knowledge synchronization for dynamic conditions (Fig. 7, column 1). Partial solutions exist [41, 40, 144], but severity scales with how distributed the system is, suggesting knowledge management should be treated as a first-class system component.

Pattern 2: Reported gains are benchmarked against weak baselines. Across the domains surveyed above, most studies compare ToSC systems against traditional schemes (JPEG, raw transmission) rather than against state-of-the-art learned compression (e.g., neural codecs, VVC). This inflates the apparent advantage of ToSC. A more honest comparison would benchmark against the best available compression method and isolate the additional gain attributable to task-awareness; the severity of this effect nonetheless varies by domain (Fig. 7), being most pronounced where no strong learned baseline is yet established.

Pattern 3: Safety-critical domains expose the brittleness trade-off. Autonomous driving, healthcare, robotics, and industrial monitoring all reveal the same tension: ToSC's efficiency gains come from discarding information deemed task-irrelevant, but the definition of "relevant" is fixed at training time. In domains where the cost of missing information is catastrophic (misdiagnosed pathology, undetected pedestrian, unsafe robot actuation near a human collaborator, unrecognized equipment failure), this brittleness is unacceptable without formal guarantees or fallback mechanisms.

Pattern 4: Evaluation methodology is inconsistent across domains. There is no standardized protocol for evaluating ToSC systems. Different domains use different metrics, channel models, compression baselines, and definitions of "task success." This makes cross-domain comparison impossible and hinders the identification of generalizable design principles. The community would benefit from standardized benchmark suites (cf. the role of ImageNet in computer vision or GLUE in NLP) that enable fair comparison across ToSC implementations. Fig. 7 maps the severity of these four patterns across all ten application domains, summarizing that knowledge dependency is the most pervasive weakness across the surveyed domains.

Application Domain	KB Dependency	Weak Baselines	Safety Brittleness	Eval Gap
Vehicular	■	■	■	◧
Satellite / UAV	■	◧	◧	◧
Surveillance	■	◧	◧	■
IoT / IIoT	■	◧	■	◧
Healthcare	■	◧	■	■
Robotics	■	◧	■	◧
Digital Twin	■	◧	◧	◧
Metaverse / XR	■	◧	□	◧
Text SC	■	■	□	◧
Speech SC	■	◧	□	◧

Legend: ■ Severe ◧ Moderate □ Minimal

Figure 7: Cross-domain weakness matrix mapping four recurring weakness patterns across ten SC/ToSC application domains. Columns represent KB dependency, weak baselines, safety-critical brittleness, and evaluation methodology gaps. Cell glyphs indicate severity (filled square = high/severe, half-filled = moderate, empty = low/minimal). Severities are the authors' qualitative synthesis of the per-domain critical assessments in Section 5, not an independent measurement; rows follow the domain order of Table 5. The uniformly severe first column reflects that knowledge dependency (an explicit shared knowledge base where one is maintained, or the implicit parametric knowledge of text- and speech-SC models) is the most pervasive weakness across the surveyed domains.

6. Research Trends in ToSC

The trends below respond to weaknesses the preceding sections exposed: the surveyed systems largely assume a fixed task, an aligned and static knowledge base, and a benign channel, and each assumption fails somewhere in deployment. Supplementary Table S4 summarizes these research areas.

6.1. Model Integration and Semantic Inference

6.1.1. Integration of ToSC Systems with LLMs and LMMs

The integration of LLMs and LMMs enables ToSC systems to dynamically reason about task requirements, overcoming the fixed task definitions that constrain traditional approaches [57]. Foundation models serve three roles: as semantic codecs that transform data across modalities (e.g., image-to-text), as a source of implicit knowledge for task decomposition, and as intelligence engines in distributed systems. Yang et al. [145] proposed M-GSC, where an LLM decomposes complex tasks and specifies standardized semantic representations, enabling scalable multi-user operation. Li et al. [146] applied goal-oriented keyframe selection for wireless video, reporting PSNR gains of 33–38% over deep-JSCC video baselines (DeepWiVe and DVST). Ribouh et al. [147] demonstrated extreme compression (a compression ratio of about 4250×, reducing one image to a 1.4 Kb textual description) by converting images to textual descriptions via a vision-language model, with a generative model synthesizing images at the receiver, trading pixel-level fidelity for the large reduction in transmitted data. Du et al. [72] explored a split architecture with a lightweight vehicular encoder and cloud-based LMM for VQA. Salehi et al. [75] instead coupled knowledge-graph extraction with LLM-based encoding and decoding for text, with the explicitly transmitted graph serving as the shared knowledge base and the model's parametric knowledge backing it; this reduced the transmitted text volume by 30%

while reaching 84% semantic similarity, against 63% for a deep-learning DeepSC baseline. More recent LMM-based designs push further on multimodality and real-world validation: Yao et al. [148] unify generation and comprehension within a single transmission by replacing the visual decoder with a multimodal language model that captions the received content without a separate text stream, while Ding et al. [149] report a foundation-model-aided video SC system evaluated over 3GPP-standard channels in simulation and validated over the air on a software-defined-radio testbed, one of the few prototype-level rather than simulation-only evaluations.

This trend loosens ToSC's dependence on task-specific models in favor of knowledge-driven designs [69]. However, LLM-based codecs introduce new failure modes: hallucinated descriptions that diverge from source content, computational costs that conflict with low-latency requirements, and poorly characterized conditions for when cross-modal transformation preserves task-relevant information. These prominently reported large-model and generative figures warrant the same methodological caution as the domain results of Section 5: each is typically a single-study, single-dataset result, and absolute compression ratios (e.g., the ~4250× of Ribouh et al.) are measured against raw or reconstruction baselines rather than a state-of-the-art learned codec (the weak-baseline caveat, Pattern 2).

The role a foundation model plays in ToSC is clarified by the three-layer knowledge taxonomy introduced earlier: a model's parameters constitute *implicit (parametric) knowledge* rather than an addressable knowledge base, so they can substitute for or back an explicit KB but cannot be queried or synchronized like one. For large-model ToSC, this framing carries practical consequences. Keeping parametric knowledge up to date requires either costly retraining or RAG; RAG couples the model to an external, updatable store of dynamic context, gaining weight-free knowledge refresh at the cost of retrieval latency and a synchronization dependency. The nearest surveyed instantiation is the knowledge-graph-plus-LLM scheme of Salehi et al. [75], which transmits an explicit graph as the shared KB while the model's parametric knowledge backs it; it is not RAG in the strict sense, since it carries structured knowledge in the message rather than querying an external index at inference. A full retrieval-augmented SC pipeline, in which the encoder or decoder queries an updatable external store in real time, is to our knowledge anticipated rather than empirically demonstrated for semantic links, mirroring the prompt-injection status noted in Section 6.4. Because parametric knowledge is implicit, the transmitter and receiver must also hold *consistent* models: differing model versions or fine-tuning yield divergent decodings of the same semantic representation, a silent failure mode analogous to knowledge-base misalignment. Most limiting for deployment, foundation models impose severe edge resource overhead: frontier foundation models now commonly contain on the order of trillions of parameters, far exceeding edge-node GPU memory, with prohibitive inference latency if invoked on every sample. To address this overhead, Zhang et al. [150] propose a cloud–edge model-evolution scheme that incrementally updates a small edge feature extractor via knowledge distillation from a cloud-resident foundation model, preserving low inference latency at a modest accuracy cost. Sun et al. [151] likewise target edge deployment with a large-AI-model agent that aligns multimodal data and plans communication policies from natural-language intent, reporting a concrete compute budget on the order of 9.4 TFLOPs (dominated by the language model and driven chiefly by input length, and met within a single consumer-grade GPU such as an RTX 4090), together with a 33% reduction in communication delay, while flagging the limited planning capacity of resource-constrained edge models. Islam and Chang [152] make the cloud-versus-edge trade-off explicit, pairing a cloud-hosted, API-based large model that attains higher semantic fidelity with a compact on-device model that removes the external API dependency at a measurable fidelity cost.

Together these results make the central deployment tension concrete for latency-sensitive settings such as vehicular networks and industrial IoT: the broad generalization that motivates foundation models strains the compute, memory, and latency budgets of edge devices, though the schemes above (cloud–edge distillation, input-length-aware edge agents) show the tension is real but tractable. These edge-generic budgets must be read against the far tighter envelopes of the latency-critical domains: vehicular V2X messaging operates under millisecond-scale end-to-end latency budgets on embedded-class automotive compute, and industrial IoT control loops run within tight low-power, limited-memory envelopes. In these regimes, even a distilled edge agent must be aggressively compressed, so the cloud–edge distillation and input-length-aware offloading schemes above are prerequisites rather than refinements for these settings. The mechanisms for closing this gap are the same ones that already shrank earlier SC models toward IoT-class budgets: pruning and quantization (L-DeepSC compressed a Transformer text encoder from 12.3 to 1.28 MB [113]), knowledge distillation to a lightweight student [153] including the cloud–edge distillation of Zhang et al. [150], and lightweight-encoder designs with fast on-device training for resource-constrained terminals [154]. Applied to a trillion-parameter foundation model, reaching the megabyte-scale memory and millisecond-scale latency envelope of automotive and industrial-IoT edge nodes would demand reductions of several orders of magnitude.

However, deployment-oriented quantities are reported only sporadically across the large-model ToSC systems surveyed here: Sun et al. give a compute budget and Zhang et al. an order-of-magnitude parameter count, while on-device latency, memory footprint, and energy per inference go unreported for the keyframe-selection, extreme-compression, split-VQA, and knowledge-graph schemes. Moreover, the budgets that do exist are edge-generic: the surveyed large-model ToSC literature reports no vehicular-network- or industrial-IoT-specific inference-latency, compute, or storage figures. Even the one explicitly vehicular scheme, the split-VQA encoder of Du et al. [72], reports only workstation-GPU response times and FLOPs rather than vehicle-grade deployment budgets, so per-domain deployment envelopes for these latency-critical settings remain an open measurement gap rather than a settled quantity, precisely the gap the standardized-evaluation argument of Section 4 targets.

6.1.2. Intent-based ToSC Systems

Intent-based systems convey the underlying purpose of a communication exchange itself, one abstraction above the task-relevant features that earlier ToSC transmits [155]. Thomas et al. [156] integrated neuro-symbolic AI (GFlowNets) to learn causal data structures, enabling nodes to reason about intent and redefining reliability based on meaning preservation. Liu et al. [153] addressed deployment constraints via knowledge distillation, transferring large-model capabilities to lightweight student models with channel-adaptive modules. Aceta et al. [157] applied this approach to industrial human-machine interaction, using ontology-driven intent interpretation to generate executable commands without large training datasets.

Intent-based communication pushes ToSC toward true Level C operation by transmitting purpose rather than content [15]. The practical limitation is that intent inference requires causal reasoning capabilities that current neuro-symbolic systems provide only for narrow domains; scaling to open-ended, multi-party interactions remains an open challenge.

6.1.3. Explainable ToSC Systems

Explainable ToSC replaces opaque feature vectors with structured, human-interpretable representations [135]. Three approaches span different modalities: Ma et al. [71] use a β-variational autoencoder to disentangle features into semantically interpretable components with selective transmission; Liu et al. [135] represent text semantics as subject-relation-object triplets with relevance-based filtering; and Sun et al. [73] extract scene graphs of objects and relationships for visual tasks (DeepSC-SG).

Explainability is essential for trust in safety-critical domains [13], but a fundamental tension persists: structured representations (scene graphs, triplets) impose a discrete bottleneck that may discard continuous semantic nuances captured by dense latent vectors, precisely the nuances that determine performance in fine-grained tasks such as medical image analysis.

6.1.4. Generative and Diffusion-Model-Based Semantic Communication

A shift is underway in semantic communication from *reconstruction-based* to *generation-based* decoding: rather than recovering a faithful replica of the transmitted signal, the receiver uses a generative model to synthesize an output that is perceptually convincing and task-appropriate, conditioned on a compact semantic description transmitted over the channel. This fundamentally changes what is preserved across the link, from source-level fidelity to semantic-level consistency.

Grassucci et al. [158] initiated this line of research by introducing a diffusion-model-based semantic communication framework in which the transmitter extracts a compact semantic map and the receiver employs a conditional diffusion process to generate high-quality reconstructions from minimal transmitted information. Yang et al. [159] proposed a diffusion-based JSCC scheme with perceptual optimization, demonstrating that diffusion decoders can achieve superior perceptual quality compared to conventional JSCC approaches, particularly at very low channel bandwidth ratios. Wu et al. [160] took a complementary physical-layer approach, introducing a channel denoising diffusion model that operates after channel equalization to learn the channel-input distribution and iteratively remove residual channel noise; combined with a JSCC encoder, it lowers reconstruction error and improves robustness across both Gaussian-noise and fading channels at varying signal-to-noise ratios. More recent generative designs couple the decoder to explicit knowledge and to new modalities: Fang et al. [161] pair a vision-language large model that extracts a concise cross-modal description with a diffusion reconstructor guided by a shared semantic knowledge base, while

Ma et al. [162] extend generative decoding to arbitrary-length video through a multimodal-large-model keyframe extractor, reaching channel-bandwidth ratios on the order of 10^{-4}–10^{-3}.

From a theoretical standpoint, generative decoding is closely connected to the rate-distortion-perception trade-off formalized by Blau and Michaeli [81], which establishes that perfect perceptual quality (i.e., the reconstructed distribution matching the source distribution) requires a higher distortion than the classical rate-distortion bound. Diffusion-based SC systems operate precisely in this regime: they sacrifice point-wise reconstruction accuracy in exchange for outputs that are distributionally faithful to the source, achieving high perceptual quality at extreme compression ratios. This reshapes the boundary between SC and ToSC: since the generated output preserves meaning and utility without preserving the source signal itself, we position generative SC not as a paradigm separate from either but as a *decoding strategy* (generation-based rather than reconstruction-based) that spans both. Where the output is judged by perceptual fidelity it serves general SC (Level B), and where it is judged by downstream task success it serves ToSC (Level C), with the rate-distortion-perception trade-off governing the balance. A broader catalog of generative diffusion models across wireless sensing, transmission, and applications is given by Fan et al. [24], and a dedicated survey of generative-AI-enabled SC (organized by generative architecture, communication modality, and application task) by Liang and Li [25].

Practical deployment nonetheless lags. Diffusion models impose substantial computational cost through their iterative sampling process, which conflicts with the low-latency requirements of many ToSC applications. Controllability is a second obstacle: ensuring that the generated output faithfully reflects the transmitted semantic description, rather than hallucinating plausible but incorrect content, requires careful conditioning mechanisms that are not yet fully understood. Formal information-theoretic guarantees are largely absent as well: unlike classical rate-distortion theory, no established framework bounds the task-performance loss incurred by generation-based decoding under arbitrary channel conditions.

6.2. *System-level Architecture*

6.2.1. *Multi-user ToSC Systems*

Multi-user ToSC introduces challenges in coordinating diverse objectives, classified as single-modal (managing interference across separate tasks) or multimodal (fusing complementary data for shared objectives) [163]. Xie et al. [12] introduced MU-DeepSC for multimodal VQA, fusing image and text features via a MAC network with jointly optimized transceivers, later extending this to a unified Transformer framework supporting both modes [54]. A related thread serves multiple *tasks* rather than multiple users from a shared encoder: Lyu et al. [52] unified coding-rate-reduction maximization, which yields discriminative features for classification, with mean-square-error minimization for image recovery in a single deep JSCC objective, and added a gated network that adaptively prunes transmitted features according to channel conditions. For resource-constrained deployment, Peng et al. [154] combined federated learning, knowledge distillation, and curriculum learning (Fed-CL-KD). Federated learning more broadly bridges single-node ToSC and distributed network AI, enabling collaborative semantic model training without aggregating raw data [164, 6].

6.2.2. *Cooperative Agents in ToSC*

Cooperative multi-agent ToSC enables agents to share distributed observations while exploiting task correlations [165, 166]. Razlighi et al. [165] proposed a split semantic encoder with a common unit for shared features and specific units for task-specific messages, later extending this to distributed sources with partial observations [166]. He et al. [167] used deep reinforcement learning to jointly learn what, when, and to whom to transmit, co-optimizing communication protocols with agent policies. A fundamental limitation is that cooperation can be counterproductive when tasks are weakly correlated: the coordination overhead may exceed collaboration gains, particularly under scarce communication resources.

6.2.3. *ToSC in Edge Computing*

Integrating ToSC with mobile edge computing (MEC) transforms semantic transmission into a joint communication-computation optimization problem, where edge devices transmit only task-essential semantic information [107]. Cang et al. [168] jointly optimized the semantic extraction factor, transmit power, and computation resources to minimize task execution delay. Guo et al. [169] introduced DTCN, where edge servers act as intelligent semantic relays that fuse device-transmitted features with their own background knowledge before forwarding. Shao et al. [109] provided theoretical grounding by deriving deterministic information-bottleneck bounds on how much video data must be delivered over the wireless channel to satisfy edge inference accuracy.

6.3. Physical Layer and Infrastructure

6.3.1. Task-Aware Encoder and Decoder Design

A growing body of work designs adaptive encoder-decoders that jointly optimize semantic prioritization with channel state awareness. Wang et al. [70] proposed I-JSCC, where the encoder scores each feature's task importance and a higher-level mechanism determines per-user transmission ratios based on both importance and channel conditions. Lyu et al.'s channel-adaptive gated network (Section 6.2.1) applies the same principle, dynamically adjusting data rate to maintain task performance [52]. Chen et al. [66] employ a masked auto-encoder that scores data patches by task importance and then selects patches for transmission based on real-time CSI. At the physical layer, Guo and Yang [170] designed MIMO precoders directly from semantic gradients rather than traditional channel metrics. A complementary robustness-oriented design pairs semantic-aware masking with a discrete codebook to harden the task-oriented link [171].

6.3.2. Semantic-Aware ToSC Systems

Semantic-aware ToSC systems explicitly evaluate the importance of each semantic feature for the downstream task and allocate wireless resources accordingly, via three complementary approaches to quantifying feature importance [172]. *Objective importance scoring* estimates each feature's contribution to task performance and transmits only the most critical ones; Wang et al. [70] introduced a metric for task-oriented semantic spectral efficiency (TOSSE) and used reinforcement learning to maximize it. *Application-specific ranking* maps importance to domain constraints: the GSAR framework [64] ranks avatar skeleton nodes by their importance to pose recovery and assigns high-priority nodes to more reliable subchannels, while Gholami et al. [173] allow flexible semantic similarity ranges that adapt compression per user. *Subjective utility models* incorporate human decision biases; Vaidanis et al. [174] used cumulative prospect theory to link feature importance to an agent's risk perception. A complementary line empirically models the semantic-loss function itself, quantifying how compression and channel conditions degrade downstream task accuracy [175].

6.3.3. Reconfigurable Intelligent Surface Assisted ToSC Systems

Reconfigurable intelligent surfaces (RIS) address a fundamental limitation of ToSC: semantic-level efficiency gains can be negated by poor physical channel quality, and joint designs that optimize source beamforming and RIS phase shifts to form favorable channels for the semantic link directly target this failure mode [176]. However, the key innovation for ToSC is optimizing RIS phase shifts based on the semantic importance of data streams, not merely channel state information [177]. Shi et al. [136] demonstrated that RIS-assisted text SC achieves higher BLEU scores in low-SNR conditions and provides robustness to channel estimation errors. Jiang et al. [178] introduced an RL-based controller that jointly optimizes subchannel allocation and RIS phase shifts based on both user needs and semantic priorities, achieving superior task performance compared to channel-only optimization. Related surface-assisted designs jointly optimize the surface configuration with semantic objectives, including stacked intelligent metasurfaces for task-oriented links [179] and a RIS-aided optimization strategy for semantic communication [180]. Mao et al. [181] broaden this surface-assisted design to joint sensing, communication, computing, and control, using RIS and semantic awareness to minimize closed-loop execution latency under a minimum-information-entropy (control-quality) constraint for industrial IoT. This dual optimization of physical channel characteristics and semantic priorities distinguishes ToSC applications of RIS from conventional signal-strength-oriented approaches [182].

6.4. Security and Privacy Against Semantic Attacks

Because ToSC systems are built on learned semantic encoders, decoders, and shared knowledge bases, their security surface differs fundamentally from that of bit-level systems: the attack target is no longer the raw payload but the AI models and the semantic representations they exchange. Transmitting only task-relevant features reduces, but does not eliminate, the exposure of source information: discarding task-irrelevant content shrinks the attack surface, yet the model-inversion attacks examined below show that recognizable source content can still be reconstructed from the transmitted features alone. Moreover, the same learned components expose AI-native vulnerabilities at every stage of the pipeline: inference-time eavesdropping and perturbation, training-time poisoning of the shared model, and large-model-specific manipulation [155]. We group the most extensively studied threats into four classes with their corresponding defenses, and then flag further AI-native risks that remain largely unexamined for semantic links.

Model inversion and privacy leakage (inference time). An eavesdropper who intercepts transmitted semantic symbols can invert the encoder to reconstruct the source. Chen et al. [183] formalized this as the *model inversion eavesdropping attack*, reconstructing recognizable images under both white-box (encoder known) and black-box (an inverse network trained from intercepted signal–image pairs) settings and across channel SNRs, and proposed a lightweight defense based on random permutation and substitution of transmitted features. However, their ablation shows that random permutation alone is insufficient: the black-box attacker, whose inverse network is trained on intercepted signal–image pairs, still reconstructs visually recognizable images under permutation-only obfuscation, and it is the combination with random substitution that renders the eavesdropped images unrecognizable under both attack types, so the protection ultimately rests on the secrecy and reliable delivery of the per-transmission permutation–substitution parameters rather than on either transformation alone.

Defenses against inversion broadly follow two routes. The first is representation disentanglement: Wang et al. [184] proposed IBAL, which uses the information-bottleneck principle (Eq. 2) with adversarial training to extract minimal task-sufficient representations that maximize adversary reconstruction distortion; Erak et al. [185] introduced CLAD, using contrastive and adversarial training to enforce statistical independence between task-relevant and private features; and Erdemir et al. [186] studied privacy-aware communication over a wiretap channel using a variational-autoencoder-based JSCC, showing through simulation that the learned encoder preserves reconstruction quality at the legitimate receiver while confusing a passive eavesdropper about a sensitive source attribute. The second is feature obfuscation: Xu et al. [187] proposed FSSC, which dynamically encrypts features during semantic extraction and transmission to block reconstruction from intercepted signals, while a prototype adversarial alignment module simultaneously hardens image tasks against adversarial perturbations. Yet these representation- and obfuscation-based defenses are bounded by limits made explicit below. IBAL and Erdemir et al.'s encoder both degrade only the eavesdropper they were trained against: a surrogate whose reconstruction distortion IBAL maximizes, and a passive listener whose sensitive-attribute leakage the variational scheme minimizes. A stronger or differently structured real eavesdropper can reduce the protected margin; adversarial robustness (below) carries the same threat-model dependence. Moreover, CLAD's enforced statistical independence between task-relevant and private features is attainable only insofar as the private attribute is uncorrelated with the task variable: any task-predictive component of a sensitive attribute cannot be removed without conceding task performance, leaving an irreducible privacy–utility floor. And the feature encryption of FSSC, like the permutation–substitution scheme above, ultimately reduces to the secrecy and reliable delivery of its per-transmission keys rather than to the obfuscation itself.

Data poisoning and backdoor attacks (training time). Because the shared semantic model is learned, often collaboratively, an adversary who corrupts the training data or model updates can implant lasting malicious behavior. In the task-oriented multi-user setting, Peng et al. [188] showed that malicious participants can degrade a collaboratively trained semantic decoder through label-flipping, clean-label, and semantic-noise-injection poisoning. They further proposed DPAD-MUSC, an adversarial-reinforcement-learning attack–defense game that detects and excludes poisoned samples, recovering on the order of 15–25% task accuracy relative to no defense, depending on the attack type (over 15% under clean-label, over 20% under label-flipping, and over 25% under semantic-noise-injection attacks). Its scope is the limitation: a detection-and-exclusion scheme confined to data poisoning, it shares the adaptive-evasion coupling flagged below for the Wei et al. detector and does not address the model-update or backdoor attacks treated next.

Backdoor attacks are subtler still: Yang et al. [189] demonstrated a *polymorphic* backdoor (SemBugger) in which intensity-graded triggers, injected by poisoning, make a single trigger family produce multiple distinct malicious outputs while preserving benign fidelity (attack success rate above 90% at under 2 dB PSNR degradation); critically, they also derived a certified *semantic-smoothing* defense (which Yang et al. report as the first certified-robustness result for SC backdoors) that suppresses triggers with a provable lower bound on efficacy. However, as with any randomized-smoothing scheme, this certified radius is governed by the injected noise level. In the reported evaluation this clean-transmission cost is small: the defense drives attack success below 1% at only about 0.2–0.4 dB of added PSNR distortion. The operative limitation is therefore not fidelity loss but adaptivity: the guarantee weakens once an adaptive adversary drives the trigger strength beyond the certified region, the very regime such defenses bound but cannot eliminate. On the detection side, Wei et al. [190] flagged poisoned samples by their deviation in semantic feature space using a threshold-based similarity test that, unlike neuron pruning or pair-constrained schemes, neither modifies the model nor restricts the data format. That test presumes the poison induces a detectable feature-space deviation; an adaptive poisoner optimizing for feature-space stealth could hold the deviation below the threshold, so detection and evasion remain locked in an escalating cycle.

Adversarial examples and robustness (inference time). Small, deliberately crafted perturbations to the input or the transmitted signal can flip a downstream decision without perceptibly altering the source. Such perturbations span two attack surfaces: *input-space* changes applied before encoding, and *over-the-air* signal-domain perturbations crafted to survive the wireless channel, the latter studied extensively for deep-learning-based wireless and NextG links [63, 191, 192, 193]. The dominant defense is adversarial training (exposing the model to worst-case perturbations during learning), which Wang et al. [184] instantiate as adversarial learning that trains the encoder to fool a model-inversion adversary by maximizing its reconstruction distortion, and Xu et al. [187] fold into their obfuscation framework. The same primitive doubles as a robustness mechanism against non-adversarial channel distribution shift: Wei et al. [194] used projected-gradient adversarial training to mimic diverse channel distortions, improving classification accuracy by more than 10% under unseen channel states without increasing model complexity. These gains are nonetheless bounded by the threat model seen during training: adversarial training hardens the model only against perturbations resembling those it was optimized against, so an adaptive attacker using a larger budget or an unmodeled perturbation geometry can still induce errors. The defense raises the attacker's cost rather than eliminating the vulnerability. This convergence, with adversarial training serving both security and channel robustness, illustrates how AI-native defenses can be co-designed with the communication objective rather than added on afterward.

Large-model-specific threats. As Section 6.1.1 discusses, foundation models increasingly serve as semantic codecs and reasoning engines, importing the security vulnerabilities of large models into the communication link. Hoang and Le [195] showed that a backdoor embedded during pre-training of an LLM-based (BERT/RoBERTa) semantic transceiver *persists through downstream fine-tuning* and resists model pruning, by mapping triggered inputs to a predefined output representation, a threat with no analogue in fixed-codebook SC. Beyond backdoors, prompt-injection and instruction-manipulation attacks, in which transmitted content is parsed by the receiver model as instructions rather than as data, are a recognized large-model risk still largely unexamined in the SC setting. They are most acute for the agentic ToSC systems of Section 6.1.1: an agent that plans communication policies from natural-language intent [151], together with any tool use it triggers, exposes an instruction channel an attacker can hijack. Candidate mitigations (isolating instruction from data, authenticating message provenance, and policy-filtering agent actions) are established for general large-model systems but not yet adapted to semantic links; this threat is therefore anticipated rather than empirically demonstrated, and remains an urgent open problem. These large-model threats are especially consequential because the same parametric knowledge is reused across tasks and users, so a single compromised model can propagate errors system-wide.

Further AI-native threats (largely unexamined). The four classes above are those with demonstrated SC/ToSC instantiations and defenses; several AI-native risks documented for learned models more broadly remain open for semantic links and should not be read as out of scope. We use *model inversion* above in the communication-specific sense of inverting the encoder to reconstruct the transmitted *input*; distinct from it are the classical attribute model-inversion attack, which recovers training-data attributes from model access [196], and membership-inference attacks, which test whether a record was used in training [197], both pertinent once a shared semantic model or knowledge base is exposed. Model-extraction (stealing) attacks can clone the encoder or KB from query access [198], against which credential-based access control that renders a stolen semantic model low-fidelity for unauthorized users is an early defense [199]; gradient-inversion attacks recover inputs from shared gradients [200], and the closely related embedding-inversion attacks recover them from the intermediate representations that semantic systems transmit; and availability attacks (denial of service, and sponge inputs that inflate inference energy and latency [201]) target the real-time guarantees ToSC depends on. Most SC-specific of all, the shared knowledge base or retrieval-augmented store that Section 6.1.1 couples to large-model ToSC is itself an *integrity* target distinct from training-time model poisoning: an adversary who poisons or tampers with entries in the shared KB or the retrieval index can steer decoding system-wide at inference time without ever modifying the model weights, a threat with no analogue in bit-level systems and directly implied by this survey's own finding that knowledge dependency is the dominant cross-domain bottleneck (Section 5). Beyond integrity, the store's confidentiality is also at stake: Hu et al. [202] demonstrate a reconstruction attack in which an adversarial large-language-model API provider recovers nearly 90% of a retrieval-augmented knowledge base through adaptive embedding-space queries. Authentication and provenance for KB entries, together with integrity verification of the retrieval index, are the natural defenses but remain unaddressed for semantic links. Watermarking and provenance for transmitted semantic models, needed to attribute a leaked or tampered model, are still at an early stage. These are named open problems (Section 7), not problems already addressed.

These vulnerabilities are most damaging precisely in the safety-critical domains of Pattern 3 (Section 5), where ToSC is most attractive and where a reconstructed private attribute, a poisoned task decision, or a backdoored large model can cause physical harm rather than mere data loss. Two lessons cut across all four threat classes. Security in ToSC is inseparable from the learning pipeline: defenses must protect the model and its shared knowledge, not merely encrypt the channel. And every defense surveyed trades against task utility or computational cost, with formal, verifiable guarantees still rare; certified defenses such as that of Yang et al. [189] are the exception rather than the rule. Because several of these threats can co-occur in practice, recent designs also compose multiple specialized defenses within a single system, for instance a mixture-of-experts that activates different security experts according to each node's declared security requirements [203]. Standardized adversarial benchmarks and verifiable, certifiable defense mechanisms are therefore a priority research direction (Section 7).

7. Challenges and Future Research Opportunities

The research trends in Section 6 and the cross-domain analysis in Section 5 reveal a field that has made rapid progress in demonstrating ToSC's potential but has not yet resolved the foundational issues preventing deployment at scale. The remaining obstacles are of three kinds, and they are not equally mature: the theory has gaps, the algorithms face known but unsolved barriers, and real-world deployment has barely been attempted. For each, we identify the specific research trends it affects and propose concrete research directions, prioritized by their potential to unblock multiple downstream problems simultaneously.

7.1. Challenges in ToSC

7.1.1. Challenges in Theory

The theoretical foundations of ToSC remain incomplete in ways that directly constrain the research trends discussed in Section 6. We identify four foundational gaps, ordered by their downstream impact:

- Definition and Quantification of SI: Shannon's classical framework [31] treats all bits as equally valuable, but the IB formulation (Eq. 2) reveals that the *relevant* information content of a source X with respect to a task T is $I(X;T)$, which can be far smaller than $H(X)$. No universally accepted measure of this task-relevant semantic content exists. Existing proposals, including logical informativeness from SIT [33], truthfulness-weighted informativeness from SSIT [34], and mutual-information-based relevance from the IB framework [26], each capture different facets but have not been unified into a single, computable metric that applies across modalities and tasks. A viable semantic entropy measure must satisfy three requirements: (1) it must reduce to Shannon entropy when the task is full reconstruction ($T = X$), (2) it must decrease monotonically as task-irrelevant information is discarded, and (3) it must be estimable from finite samples via variational bounds [65]. Developing and validating such a measure is the deepest open problem here: until one exists, task-relevant information can be bounded (Eq. 2) but not directly measured.
- Modeling Semantic Noise: Semantic noise arises not from channel impairments but from mismatches in interpretation between sender and receiver. Bao et al.'s semantic channel model (Section 2) formalized one source of this noise: when the transmitter's and receiver's KBs diverge ($K_s \neq K_d$), the receiver's semantic-interpretation term $\overline{H}_s(Y)$ falls (the receiver recovers less model-theoretic meaning from the same channel output under its mismatched knowledge base), so the semantic channel capacity C_{sem} degrades independently of the physical channel. Güler et al. [39] further showed that under asymmetric information, the achievable meaning recovery is bounded by the Bayesian game equilibrium, which may be far below the semantic capacity achievable with aligned KBs. Yet current theoretical models treat semantic noise either as a fixed KB offset or as an additive perturbation analogous to channel noise; neither captures the *compositional* nature of real-world semantic mismatches, where a single misaligned concept (e.g., confusing "high vibration from tool wear" with "vibration from an adjacent press" in an IIoT setting) can cascade into task failure. Recent work begins to move in this direction: Fayaz et al. [204] decompose semantic noise into source-, channel-, and receiver-side

components and show that source- and receiver-side noise can dominate task performance even when channel-induced semantic errors are negligible. New models must represent semantic noise as a structured, concept-level perturbation that interacts with the task variable T, not merely the source X.

- Semantic Rate-Distortion Theory: Classical rate-distortion theory characterizes the minimum rate needed to reconstruct a source within a fidelity constraint, but ToSC requires characterizing the minimum rate needed to *perform a task* within a performance constraint, a problem of a different kind when the task requires strictly less information than the source (i.e., $I(X;T) < H(X)$, the regime where SC and ToSC diverge, as analyzed in Section 3.6). This problem has deep roots in information theory. The remote source coding framework, originating with Dobrushin and Tsybakov [205] and Witsenhausen [206], established rate-distortion bounds for compressing a source observed through a noisy mapping, precisely the setting of task-oriented compression where the encoder observes raw data X but cares only about a downstream variable T. The CEO problem [207] extends this to distributed settings directly applicable to multi-agent ToSC. Stavrou and Kountouris [208] recently formalized a rate-distortion approach to goal-oriented communication, providing the closest existing theoretical treatment of what the field needs. Liu et al. [209] further formulated rate-distortion bounds when the transmitter observes only an indirect version of the semantic source, proving divergence from Shannon limits. Blau and Michaeli [81] demonstrated an additional complication: rate-distortion bounds conflict with perceptual quality metrics, meaning that minimizing reconstruction error forces divergence from true semantic perception. The remaining gap is therefore narrower than sometimes claimed: it is specifically a *KB-conditioned* extension of these known frameworks that incorporates KB alignment quality as a parameter. When KBs are perfectly aligned, the encoder can exploit shared knowledge to achieve lower rates; when KBs degrade, the effective rate-distortion bound shifts, requiring higher rates to compensate for semantic ambiguity. Developing such KB-dependent bounds (cf. Priority 1 in the research roadmap below) would provide the first principled benchmarks for ToSC system design.
- Linking Semantics to Task Performance: The IB framework provides the formal structure for this link, since $I(Z;T)$ measures how much task-relevant information the representation Z preserves, but translating this into closed-form performance predictions for practical systems remains open. The concrete divergence evidence from Section 3.6 (Cases 1–3) demonstrates that the relationship between semantic representation quality and task performance is non-trivial. Kang et al. [13] showed that task-oriented and reconstruction-oriented encoders *diverge in what they preserve*, with the personalized task-oriented encoder transmitting only the user-relevant scene-graph subgraph (on the order of a 64% communication-cost reduction) while a reconstruction objective must retain the whole image. Lyu et al. [52] reported a qualitative accuracy–PSNR trade-off (their Fig. 10) on which classification accuracy and reconstruction PSNR are not simultaneously maximized. KB mismatch introduces an additional confounding factor that current models ignore: even a representation with high $I(Z;T)$ under aligned KBs may yield poor task performance when $K_s \neq K_d$, because the receiver misinterprets the semantic content; analogously, a representation with high $I(Z;T)$ on the training distribution can degrade sharply under the domain and semantic shifts encountered at deployment [210]. Establishing rigorous, KB-aware performance bounds that predict task accuracy as a function of representation quality, channel conditions, *and* KB alignment is essential for moving ToSC from empirical demonstration to principled engineering.

7.1.2. Challenges in Algorithm Development

The research trends in Section 6 demonstrate advanced algorithmic solutions, but several fundamental barriers remain that cut across all trend areas:

- Joint Source-Channel Coding: ToSC requires a joint approach where semantic content is encoded and protected against channel impairments simultaneously, often through end-to-end learning [15]. The resulting optimization problem is non-convex: it involves tuning the high-dimensional, non-linear parameters of neural networks for both the source and channel coding aspects. Kostina and Verdú [58] showed that JSCC can strictly outperform separate coding in the finite blocklength regime, yet the optimal joint mapping remains intractable for DNN-parameterized systems. A specific technical gap is the difficulty of backpropagating gradients through non-differentiable wireless channel layers: current solutions often rely on simplified, differentiable channel approximations or reinforcement learning, which may not capture the full complexity of real-world effects like deep fades or non-Gaussian interference, limiting the robustness of the learned solution.

- Robust Optimization Under Uncertainty: ToSC systems must operate reliably in real-world situations involving inherent uncertainty, including variations in wireless channel conditions, environmental interference, and, uniquely, *semantic noise*. Hu et al. [114], motivating their robust masked VQ-VAE codebook, report that conventional semantic systems are degraded by such noise: adversarial, sample-dependent and sample-independent perturbations that distort the intended meaning, arising from ambiguities in language or data representation, mismatches in contextual understanding, or discrepancies in interpretation between sender and recipient. This requires optimization techniques that maintain performance under these stochastic variations while guaranteeing the intended meaning is accurately conveyed and used despite potential semantic ambiguities.
- Achieving Robust and Resource-Aware Multi-Task Generalization: Developing ToSC systems that perform a wide range of tasks, potentially simultaneously, and still generalize to new, unknown tasks or changing environmental/network conditions requires algorithms that learn representations beneficial for multiple goals while concurrently optimizing often-conflicting objectives. Transfer across heterogeneous tasks is the hard case; for example, a semantic representation optimized for a classification task may be entirely unsuitable for a generative one, such as image reconstruction. The resource-aware dimension is partially addressed by the adaptable-compression framework of Liu et al. [101], which jointly adapts the semantic compression ratio and wireless resources across users and channel conditions, though transferring a single semantic representation across heterogeneous tasks remains largely open; versatile deployment depends on balancing these competing trade-offs.
- Lack of Unified Semantic-Related Performance Evaluation Metrics: Algorithmic development is also slowed by the lack of broadly acknowledged, unified performance metrics for reliably assessing semantic-related performance [50]. Unlike traditional communication systems, which rely on measurements such as bit error rate, ToSC needs measures that both reliably correlate with actual task success across applications and are mathematically well-founded. This absence of uniform criteria makes it difficult to objectively compare algorithmic approaches and steer the optimization process.
- Lack of Interpretability and Explainability for DL-based ToSC Systems: Many modern ToSC algorithms, especially those based on deep learning, operate as "black boxes" [71, 135]: this opacity makes it difficult to grasp how semantic features are detected, retrieved, and processed and how these learned representations contribute to task outcomes, in turn making it difficult to diagnose errors, improve algorithm design, or establish trust in these systems, especially in critical applications. Future research should emphasize algorithms that incorporate explainable AI methodologies to clarify semantic feature selection and decision-making.
- Computational Complexity and Scalability: Modern semantic processing frequently relies on deep DNNs and complex optimization techniques that require significant compute power and memory, making implementation on resource-constrained devices (IoT sensors, mobile devices), and scaling to vast, heterogeneous networks, hard on both the algorithmic and the engineering side [113, 153]. Still needed are lightweight algorithms, model compression techniques, and distributed learning frameworks that reduce computing cost while maintaining semantic fidelity and task performance.

7.1.3. Challenges for Real-World Deployment

The gap between the laboratory results reported in Sections 5–6 and operational deployment is substantial. The cross-domain patterns identified in Section 5 (knowledge dependency and brittleness most directly, with the two evaluation-side patterns addressed under Priority 2 below) manifest as concrete engineering challenges:

- Hardware and Computational Constraints: Many advanced semantic processing algorithms, particularly those based on DNNs, are computationally intensive, demanding substantial processing power and memory. Implementing them on edge devices or under limited power budgets demands hardware-friendly implementations and efficient algorithmic optimizations that cut latency and energy consumption while preserving semantic fidelity.
- Interoperability with Legacy Systems and Standardization: Existing communication infrastructures are largely built upon traditional, bit-centric protocols, so integrating a new semantic layer requires ensuring interoperability and coexistence with existing standards and hardware, via interfaces and transition strategies that allow a gradual shift toward semantic communication. The lack of standardization for SI representation, protocols, and KBs further complicates interoperability between systems from different vendors or designed for different

applications; future deployment should rely on the development of common standards and dynamic resource management frameworks.

- Environmental and Channel Variability: Real-world wireless networks are characterized by highly dynamic and unpredictable conditions, including fluctuating channel quality, varying interference levels, and user mobility, so ToSC systems that consistently extract, transmit, and interpret semantic content despite these variations typically need real-time adaptive mechanisms that continuously adjust semantic models and communication strategies to the current operational environment.
- KB Alignment and Management Problem: Deployment also faces the KB alignment and management problem: complete and accurate KBs are expensive to build, and even once established, independent updates or differing training experiences pull distributed transmitters and receivers apart, degrading performance; dynamically updating KBs across distributed systems compounds the difficulty. A related issue is the inherent inconsistency between source and destination KBs, even assuming real-time sharing: goal-oriented semantic-language frameworks assume the shared language and KB track local, time-varying goals [211], while Farshbafan et al. [40] proposed curriculum learning to actively synchronize a common language but at significant alignment overhead. As such, ToSC systems, particularly multi-user ones, must be designed to operate despite KB discrepancies, a requirement none of the surveyed designs yet meets robustly.
- Security and Privacy Concerns: Transmitting semantically rich information introduces new security vulnerabilities and privacy risks: malicious actors can manipulate the semantic content to cause system failure or inject harmful instructions, while even unmanipulated semantic features can inadvertently reveal sensitive personal or operational details. This requires privacy-preserving semantic compression that sanitizes data while retaining task-relevant utility, together with authentication and integrity protection for the semantic content itself and for the shared knowledge base and retrieval store against poisoning or tampering.

7.2. Prioritized Research Directions

Rather than enumerating all possible future directions, we distill the challenges above into five high-priority research problems, ranked by the number of downstream barriers each unblocks and informed by the cross-domain patterns identified in Section 5 and the limitations observed across the research trends in Section 6.

Priority 1 [Theoretical]: Semantic rate-distortion theory with KB-dependent bounds. This is the theoretical contribution the field most needs. A formal framework linking transmission rate, task-specific distortion (not signal-level distortion), and KB alignment quality would provide: (a) fundamental limits against which all ToSC systems can be benchmarked, (b) principled guidance for encoder design rather than empirical trial-and-error, and (c) a formal basis for the SC/ToSC trade-off analyzed in Section 3.6. Viable directions include extending the information bottleneck framework (Eq. 2) to incorporate KB mismatch as a noise source [210], and using variational methods [65] to approximate the resulting bounds for practical DNN-based systems.

Priority 2 [Engineering]: Standardized benchmarks and evaluation protocols. The cross-domain analysis in Section 5 identified inconsistent evaluation as a field-wide problem. The community urgently needs: (a) public, multi-modal datasets with both reconstruction and task annotations, enabling fair SC-vs-ToSC comparison; (b) a standardized evaluation protocol specifying channel models, compression baselines (including state-of-the-art neural codecs, not just JPEG), and reporting requirements; (c) metrics that capture cross-task generalizability, not just single-task performance; and (d) an open, hosted evaluation platform (a public leaderboard with reference implementations and held-out test channels) that reports deployment-oriented measures (inference latency, model size, energy per inference, and robustness to KB mismatch) alongside task accuracy, so that comparisons reflect practical deployability rather than benchmark accuracy in isolation. A template for this exists in adjacent fields (ImageNet, GLUE, LibriSpeech). Establishing such benchmarks and the platform to run them would accelerate reproducible research and enable the field to move beyond inflated comparisons against weak baselines.

Priority 3 [Algorithmic/Engineering]: Scalable KB management with formal consistency guarantees. Knowledge dependency emerged as the most pervasive bottleneck across the surveyed application domains. Deployment at scale requires distributed KB management systems that provide: (a) provable consistency guarantees under network partitions and asynchronous updates, drawing on techniques from distributed databases and consensus algorithms, specifically conflict-free replicated data types (CRDTs) for eventual consistency without coordination overhead, Raft-based consensus for strong consistency in small clusters, and vector clock-based versioning for detecting

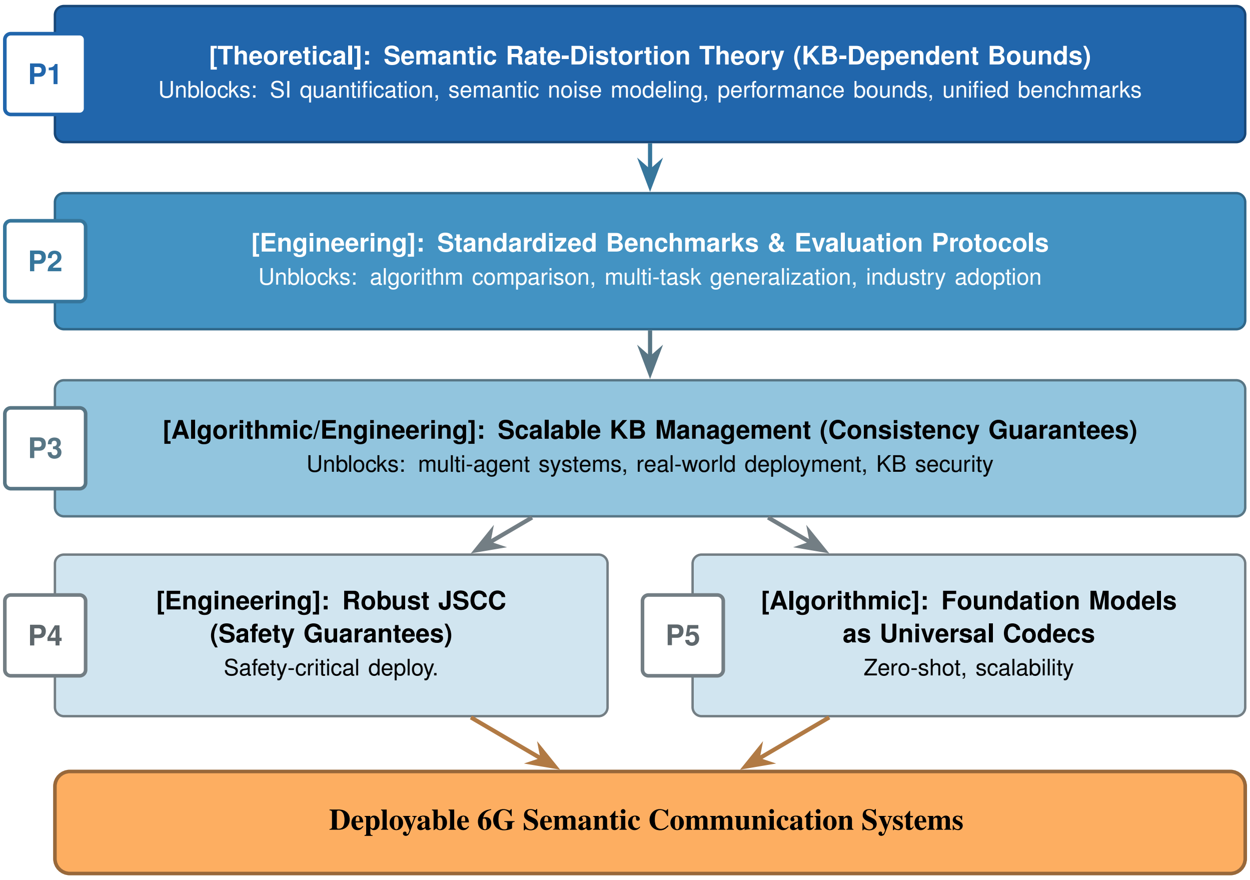

Figure 8: Prioritized research roadmap for semantic communication, structured as a top-down dependency hierarchy. Priority 1 (semantic rate-distortion theory with KB-dependent bounds) is the enabling root at the top, unblocking standardized benchmarks (Priority 2) and, via them, KB management design (Priority 3). Priority 3 in turn unblocks both robust JSCC for safety-critical deployment (Priority 4) and foundation model-based universal codecs (Priority 5). Arrows indicate enabling dependencies: solving lower-layer priorities removes barriers for higher-layer ones. The ordering reflects the analysis in Sections 5 and 7, where each priority is ranked by the number of downstream challenges it unblocks.

and resolving KB divergence across asynchronous agents; (b) graceful degradation protocols that detect KB drift (e.g., via Bayesian KB alignment monitoring) and increase transmission redundancy or fall back to SC mode; (c) lightweight KB synchronization mechanisms suitable for resource-constrained devices (IoT sensors, UAVs). Engagement with standardization bodies (ETSI ISG ZSM, O-RAN Alliance, the IEEE P1876 working group on semantic communication standards, and the 3GPP SA1 Study Item on AI/ML-assisted communication) will be essential for ensuring that KB management solutions achieve cross-vendor interoperability.

Priority 4 [Engineering]: Robust JSCC with formal safety guarantees for critical applications. The brittleness trade-off identified in safety-critical domains (Pattern 3) demands JSCC algorithms that provide: (a) formal worst-case performance bounds, not just average-case improvements; (b) graceful fallback to higher-fidelity transmission when task confidence drops below a threshold; (c) robustness to non-differentiable channel effects (deep fades, non-Gaussian interference) that current gradient-based training cannot handle; and (d) certified robustness to adversarial, poisoning, and backdoor attacks, supported by standardized adversarial benchmarks, so that the AI-native defenses surveyed in Section 6.4 can be compared and verified rather than assessed ad hoc. Hybrid model-driven and data-driven approaches offer a path toward interpretable, certifiable JSCC systems: deep unfolding [212] maps iterative optimization algorithms onto trainable neural network layers, preserving convergence guarantees while learning channel-adapted parameters; neuro-symbolic methods [156] combine learned feature extraction with rule-based reasoning to provide verifiable decision paths.

Priority 5 [Algorithmic]: Foundation models as universal semantic codecs. Large-scale foundation models (LLMs and LMMs) offer a potential solution to the KB alignment problem by substituting for an explicit, shared knowledge base. The key research challenges are: (a) efficient adaptation and distillation of foundation models for edge deployment, achieving the compression benefits demonstrated in Section 6 without the computational overhead; (b) formal analysis of when a foundation model's parametric knowledge provides sufficient alignment for ToSC tasks vs. when domain-specific fine-tuning is required; (c) security implications of relying on centralized foundation models as a substitute for shared KBs (single point of attack, model and knowledge-base/retrieval-store poisoning). A closely related challenge is the emergence of generative and diffusion-model-based decoding (Section 6.1), which fundamentally shifts the SC/ToSC boundary by replacing source-level reconstruction with semantic-level generation; extending the rate-distortion-perception trade-off [81] to formally bound task-performance loss under generative decoding remains an open theoretical problem that intersects with Priorities 1 and 4. Fig. 8 visualizes the dependency ordering among the five priorities.

8. Conclusion

This survey aimed to answer the questions that catalog-style surveys of SC leave open: when the distinction between SC and ToSC actually matters, where current systems fall short, and what the field must prioritize to move from laboratory demonstrations to deployable technology. Three cross-cutting findings emerge from our analysis, synthesizing the conditional SC/ToSC distinction of Section 3 with the four cross-domain weakness patterns of Section 5: the third finding consolidates the weak-baseline and evaluation-inconsistency patterns, while the brittleness pattern is carried forward as research Priority 4 (robust JSCC with formal safety guarantees).

First, the SC/ToSC distinction is substantive but conditional. Synthesizing the information bottleneck framework (Section 3.6) with empirical evidence from the surveyed literature, we identified that SC and ToSC diverge whenever the task requires strictly less information than the source contains, a condition that holds for most practical applications involving high-dimensional data and low-dimensional tasks. However, the distinction collapses in low-rate regimes and for tasks requiring full source fidelity (e.g., open-ended diagnostic reading of medical images), whereas for low-variance, high-value features (e.g., detecting a tiny lesion within a large irrelevant background) extreme compression can instead drive the two apart; the choice between SC and ToSC should therefore be an explicit engineering decision based on the task-information gap, not a default assumption.

Second, knowledge dependency is the field's most underestimated weakness. Nearly every application domain we surveyed depends on aligned knowledge, whether an explicit shared knowledge base or the implicit knowledge embedded in model weights, yet none has solved knowledge management for real-world conditions. The theoretical framework of Bao et al. shows that KB mismatch introduces semantic noise that degrades capacity independently of physical channel quality, and our cross-domain analysis confirms this as the most pervasive bottleneck. We further argue that ToSC systems, by discarding "redundant" information an SC receiver could otherwise use to absorb KB errors, are *plausibly more* sensitive to this problem than general SC systems (a conjecture grounded in the information-bottleneck view rather than a direct empirical comparison, whose rigorous test depends on the KB-dependent bounds and a controlled SC-vs-ToSC robustness benchmark we identify as open problems in Priorities 1–2), with direct implications for deployment strategy.

Third, the field's evaluation methodology systematically overstates ToSC's advantages: as the weak-baseline and evaluation-inconsistency patterns of Section 5 document, reported gains are typically measured against weak baselines (traditional codecs or raw transmission) under controlled channel conditions, with no standardized protocol for cross-system comparison. Establishing rigorous, standardized benchmarks is not merely optional but a prerequisite for the field's scientific credibility.

Based on these findings, we proposed five research priorities ordered by the number of downstream barriers each unblocks: semantic rate-distortion theory with KB-dependent bounds, standardized benchmarks, scalable KB management with consistency guarantees, robust JSCC with formal safety guarantees, and foundation models as universal semantic codecs. These priorities are dependency-ordered, not a wish list. SC and ToSC change what 6G communication systems optimize for: meaning and task effectiveness instead of bit fidelity. The efficiency gains documented across ten application domains suggest the practical value of this shift, although their magnitude is likely overstated by the prevalent benchmarking against weak baselines (Section 5) and remains to be confirmed against state-of-the-art neural codecs. However, realizing this potential requires the field to move beyond demonstrating what ToSC *can* do under favorable conditions and instead establish what it *guarantees* under realistic ones. The analytical framework, critical assessments, and prioritized research directions presented in this survey are intended to accelerate that transition.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used Claude (Anthropic) in order to improve the language and readability of the manuscript and to assist in editing during the revision. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] S. Dang, O. Amin, B. Shihada, M.-S. Alouini, What should 6G be?, Nat. Electron. 3 (1) (2020) 20–29. doi:10.1038/s41928-019-0355-6.

[2] W. Saad, M. Bennis, M. Chen, A vision of 6G wireless systems: Applications, trends, technologies, and open research problems, IEEE Netw. 34 (3) (2019) 134–142. doi:10.1109/mnet.001.1900287.

[3] P. Popovski, F. Chiariotti, V. Croisfelt, A. E. Kalør, I. Leyva-Mayorga, L. Marchegiani, S. R. Pandey, B. Soret, Internet of things (iot) connectivity in 6g: An interplay of time, space, intelligence, and value, arXiv preprint arXiv:2111.05811 (2021).

[4] G. Xin, P. Fan, K. B. Letaief, Semantic communication: A survey of its theoretical development, Entropy 26 (2) (2024) 102. doi:10.3390/e26020102.

[5] L.-H. Shen, K.-T. Feng, L. Hanzo, Five facets of 6g: Research challenges and opportunities, ACM Comput. Surv. 55 (11) (2023) 1–39. doi:10.1145/3571072.

[6] S. Zhang, W. Y. B. Lim, W. C. Ng, Z. Xiong, D. Niyato, X. S. Shen, C. Miao, Toward green metaverse networking: Technologies, advancements, and future directions, IEEE Netw. 37 (5) (2023) 223–232. doi:10.1109/mnet.130.2200510.

[7] E. C. Strinati, S. Barbarossa, 6G networks: Beyond Shannon towards semantic and goal-oriented communications, Comput. Netw. 190 (2021) 107930. doi:10.1016/j.comnet.2021.107930.

[8] W. Yang, H. Du, Z. Q. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. Shen, C. Miao, Semantic communications for future internet: Fundamentals, applications, and challenges, IEEE Commun. Surv. Tutor. 25 (1) (2022) 213–250. doi:10.1109/comst.2022.3223224.

[9] K. Niu, J. Dai, S. Yao, S. Wang, Z. Si, X. Qin, P. Zhang, A paradigm shift toward semantic communications, IEEE Commun. Mag. 60 (11) (2022) 113–119. doi:10.1109/mcom.001.2200099.

[10] Y. Wang, H. Han, Y. Feng, J. Zheng, B. Zhang, Semantic communication empowered 6g networks: Techniques, applications, and challenges, IEEE Access (2025). doi:10.1109/access.2025.3532797.

[11] K. Lu, Q. Zhou, R. Li, Z. Zhao, X. Chen, J. Wu, H. Zhang, Rethinking modern communication from semantic coding to semantic communication, IEEE Wireless Commun. 30 (1) (2022) 158–164. doi:10.1109/mwc.013.2100642.

[12] H. Xie, Z. Qin, G. Y. Li, Task-oriented multi-user semantic communications for VQA, IEEE Wireless Commun. Lett. 11 (3) (2021) 553–557. doi:10.1109/lwc.2021.3136045.

[13] J. Kang, H. Du, Z. Li, Z. Xiong, S. Ma, D. Niyato, Y. Li, Personalized saliency in task-oriented semantic communications: Image transmission and performance analysis, IEEE J. Sel. Areas Commun. 41 (1) (2022) 186–201. doi:10.1109/jsac.2022.3221990.

[14] T. M. Getu, G. Kaddoum, M. Bennis, A survey on goal-oriented semantic communication: Techniques, challenges, and future directions, IEEE Access (2024). doi:10.1109/access.2024.3381967.

[15] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, C.-B. Chae, Beyond transmitting bits: Context, semantics, and task-oriented communications, IEEE J. Sel. Areas Commun. 41 (1) (2023) 5–41. doi:10.1109/jsac.2022.3223408.

[16] Y. Liu, X. Wang, Z. Ning, M. Zhou, L. Guo, B. Jedari, A survey on semantic communications: technologies, solutions, applications and challenges, Digit. Commun. Netw. (2024). doi:10.1016/j.dcan.2023.05.010.

[17] Z. Lu, R. Li, K. Lu, X. Chen, E. Hossain, Z. Zhao, H. Zhang, Semantics-empowered communications: A tutorial-cum-survey, IEEE Commun. Surv. Tutor. (2023). doi:10.1109/comst.2023.3333342.

[18] D. Wheeler, B. Natarajan, Engineering semantic communication: A survey, IEEE Access 11 (2023) 13965–13995. doi:10.1109/access.2023.3243065.

[19] T. M. Getu, G. Kaddoum, M. Bennis, Making sense of meaning: A survey on metrics for semantic and goal-oriented communication, IEEE Access 11 (2023) 45456–45492. doi:10.1109/access.2023.3269848.

[20] S. Guo, Y. Wang, N. Zhang, Z. Su, T. H. Luan, Z. Tian, X. Shen, A survey on semantic communication networks: Architecture, security, and privacy, IEEE Commun. Surv. Tutor. (2024). doi:10.1109/comst.2024.3516819.

[21] L. X. Nguyen, A. D. Raha, P. S. Aung, D. Niyato, Z. Han, C. S. Hong, A contemporary survey on semantic communications: Theory of mind, generative ai, and deep joint source-channel coding, IEEE Commun. Surv. Tutor. 28 (2026) 2377–2417, arXiv:2502.16468. doi:10.1109/comst.2025.3616973.

[22] R. Meng, S. Gao, D. Fan, H. Gao, Y. Wang, X. Xu, B. Wang, S. Lv, Z. Zhang, M. Sun, et al., A survey of secure semantic communications, J. Netw. Comput. Appl. (2025) 104181doi:10.1016/j.jnca.2025.104181.

[23] R. Meng, Z. Huang, J. Yan, M. Sun, Y. Liu, C. Feng, X. Xu, Z. Zhang, S. Gao, P. Zhang, T. Q. S. Quek, Semantic radio access networks: Architecture, state-of-the-art, and future directions, IEEE Trans. Cogn. Commun. Netw. 12 (2026) 7076–7097, arXiv:2512.20917. doi:10.1109/tccn.2026.3683191.

[24] D. Fan, R. Meng, X. Xu, Y. Liu, G. Nan, C. Feng, S. Han, S. Gao, B. Xu, D. Niyato, T. Q. S. Quek, P. Zhang, Generative diffusion models for wireless networks: Fundamental, architecture, and state-of-the-art, IEEE Commun. Surv. Tutor. 28 (2026) 5632–5677, arXiv:2507.16733. doi:10.1109/comst.2026.3671110.
[25] C. Liang, D. Li, Generative AI-enabled semantic communication: State-of-the-art, applications, and the way ahead, IEEE Commun. Surv. Tutor. 28 (2026) 3976–4015. doi:10.1109/comst.2025.3649707.
[26] N. Tishby, F. C. Pereira, W. Bialek, The information bottleneck method, arXiv preprint physics/0004057 (2000).
[27] E. Beck, C. Bockelmann, A. Dekorsy, Semantic information recovery in wireless networks, Sensors 23 (14) (2023) 6347. doi:10.3390/s23146347.
[28] Z. Lyu, X. Cao, X. Song, Y. Li, J. Wang, S. Wang, Y. Cui, W. Yuan, X. Yu, G. Zhu, H. Liu, J. Xu, D. W. K. Ng, S. Cui, Sense smarter, think better: A survey on edge perception for next-generation networks, arXiv preprint arXiv:2605.18457 (2026).
[29] C. W. Morris, Foundations of the theory of signs, in: International encyclopedia of unified science, Chicago University Press, 1938, pp. 1–59.
[30] C. E. Shannon, W. Weaver, The mathematical theory of communication, University of Illinois press, 1998.
[31] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (3) (1948) 379–423.
[32] W. Weaver, Recent contributions to the mathematical theory of communication, ETC: a review of general semantics (1953) 261–281.
[33] Y. Bar-Hillel, R. Carnap, Semantic information, Br. J. Philos. Sci. 4 (14) (1953) 147–157. doi:10.1093/bjps/IV.14.147.
[34] L. Floridi, Outline of a theory of strongly semantic information, Minds Mach. 14 (2004) 197–221.
[35] S. D'Alfonso, On quantifying semantic information, Information 2 (1) (2011) 61–101. doi:10.3390/info2010061.
[36] I. Niiniluoto, Truthlikeness, Vol. 185, Springer Science & Business Media, 2012.
[37] H. Xie, Z. Qin, G. Y. Li, B.-H. Juang, Deep learning enabled semantic communication systems, IEEE Trans. Signal Process. 69 (2021) 2663–2675. doi:10.1109/tsp.2021.3071210.
[38] J. Bao, P. Basu, M. Dean, C. Partridge, A. Swami, W. Leland, J. A. Hendler, Towards a theory of semantic communication, in: Proc. 2011 IEEE Netw. Sci. Workshop, IEEE, 2011, pp. 110–117. doi:10.1109/nsw.2011.6004632.
[39] B. Güler, A. Yener, A. Swami, The semantic communication game, IEEE Trans. Cogn. Commun. Netw. 4 (4) (2018) 787–802.
[40] M. K. Farshbafan, W. Saad, M. Debbah, Curriculum learning for goal-oriented semantic communications with a common language, IEEE Trans. Commun. 71 (3) (2023) 1430–1446. doi:10.1109/tcomm.2023.3236671.
[41] F. Zhou, Y. Li, X. Zhang, Q. Wu, X. Lei, R. Q. Hu, Cognitive semantic communication systems driven by knowledge graph, in: Proc. ICC 2022-IEEE Int. Conf. Commun., IEEE, 2022, pp. 4860–4865. doi:10.1109/icc45855.2022.9838470.
[42] B. Juba, M. Sudan, Universal semantic communication I, in: Proc. 40th Annu. ACM Symp. Theory Comput., 2008, pp. 123–132.
[43] B. Juba, M. Sudan, Universal semantic communication II: A theory of goal-oriented communication, Tech. Rep. TR08-095, MIT CSAIL (2009).
URL https://people.csail.mit.edu/madhu/papers/2008/usc2.pdf
[44] O. Goldreich, B. Juba, M. Sudan, A theory of goal-oriented communication, J. ACM 59 (2) (2012) 1–65. doi:10.1145/2160158.2160161.
[45] B. Juba, S. Vempala, Semantic communication for simple goals is equivalent to on-line learning, in: Algorithmic Learning Theory: 22nd International Conference, ALT 2011, Espoo, Finland, October 5-7, 2011. Proceedings 22, Springer, 2011, pp. 277–291. doi:10.1007/978-3-642-24412-4_23.
[46] M. Kountouris, N. Pappas, Semantics-empowered communication for networked intelligent systems, IEEE Commun. Mag. 59 (6) (2021) 96–102. doi:10.1109/mcom.001.2000604.
[47] J. Shao, Y. Mao, J. Zhang, Learning task-oriented communication for edge inference: An information bottleneck approach, IEEE J. Sel. Areas Commun. 40 (1) (2021) 197–211. doi:10.1109/jsac.2021.3126087.
[48] Y. E. Sagduyu, T. Erpek, A. Yener, S. Ulukus, Will 6g be semantic communications? opportunities and challenges from task oriented and secure communications to integrated sensing, IEEE Netw. (2024). doi:10.1109/mnet.2024.3425152.
[49] H. Zhou, Y. Deng, X. Liu, N. Pappas, A. Nallanathan, Goal-oriented semantic communications for 6g networks, IEEE Internet Things Mag. 7 (5) (2024) 104–110. doi:10.1109/iotm.001.2300269.
[50] S. Barbarossa, D. Comminiello, E. Grassucci, F. Pezone, S. Sardellitti, P. Di Lorenzo, Semantic communications based on adaptive generative models and information bottleneck, IEEE Commun. Mag. 61 (11) (2023) 36–41. doi:10.1109/mcom.005.2200829.
[51] E. Bourtsoulatze, D. B. Kurka, D. Gündüz, Deep joint source-channel coding for wireless image transmission, IEEE Trans. Cogn. Commun. Netw. 5 (3) (2019) 567–579. doi:10.1109/TCCN.2019.2919300.
[52] Z. Lyu, G. Zhu, J. Xu, B. Ai, S. Cui, Semantic communications for image recovery and classification via deep joint source and channel coding, IEEE Trans. Wireless Commun. 23 (8) (2024) 8388–8404. doi:10.1109/TWC.2023.3349330.
[53] E. Eldeeb, M. Shehab, H. Alves, A multi-task oriented semantic communication framework for autonomous vehicles, arXiv preprint arXiv:2403.12997 (2024).
[54] H. Xie, Z. Qin, X. Tao, K. B. Letaief, Task-oriented multi-user semantic communications, IEEE J. Sel. Areas Commun. 40 (9) (2022) 2584–2597. doi:10.1109/jsac.2022.3191326.
[55] J. Xu, T.-Y. Tung, B. Ai, W. Chen, Y. Sun, D. Gündüz, Deep joint source-channel coding for semantic communications, IEEE Commun. Mag. 61 (11) (2023) 42–48. doi:10.1109/mcom.004.2200819.
[56] E. C. Strinati, P. Di Lorenzo, V. Sciancalepore, A. Aijaz, M. Kountouris, D. Gündüz, P. Popovski, M. Sana, P. A. Stavrou, B. Soret, et al., Goal-oriented and semantic communication in 6G AI-native networks: The 6G-GOALS approach, in: Proc. 2024 Joint Eur. Conf. Netw. Commun. & 6G Summit (EuCNC/6G Summit), IEEE, 2024, pp. 1–6. doi:10.1109/eucnc/6gsummit60053.2024.10597087.
[57] M. Chen, M. Liu, Z. Zhang, Z. Xu, L. Wang, Task-oriented semantic communication with foundation models, China Commun. 21 (7) (2024) 65–77. doi:10.23919/jcc.fa.2024-0024.202407.
[58] V. Kostina, S. Verdú, Lossy joint source-channel coding in the finite blocklength regime, IEEE Trans. Inf. Theory 59 (5) (2013) 2545–2575. doi:10.1109/isit.2012.6283533.
[59] D. B. Kurka, D. Gündüz, Bandwidth-agile image transmission with deep joint source-channel coding, IEEE Trans. Wireless Commun. 20 (12) (2021) 8081–8095. doi:10.1109/twc.2021.3090048.

[60] J. Dai, S. Wang, K. Tan, Z. Si, X. Qin, K. Niu, P. Zhang, Nonlinear transform source-channel coding for semantic communications, IEEE J. Sel. Areas Commun. 40 (8) (2022) 2300–2316. doi:10.1109/jsac.2022.3180802.

[61] D. B. Kurka, D. Gündüz, DeepJSCC-f: Deep joint source-channel coding of images with feedback, IEEE J. Sel. Areas Inf. Theory 1 (1) (2020) 178–193. doi:10.1109/JSAIT.2020.2987203.

[62] T.-Y. Tung, D. B. Kurka, M. Jankowski, D. Gündüz, DeepJSCC-Q: Channel input constrained deep joint source-channel coding, in: Proc. IEEE Int. Conf. Commun. (ICC), IEEE, 2022, pp. 3880–3885. doi:10.1109/icc45855.2022.9838671.

[63] Y. E. Sagduyu, S. Ulukus, A. Yener, Task-oriented communications for NextG: End-to-end deep learning and AI security aspects, IEEE Wireless Commun. 30 (3) (2023) 52–60. doi:10.1109/mwc.006.2200494.

[64] Z. Wang, Y. Deng, A. H. Aghvami, Goal-oriented semantic communications for avatar-centric augmented reality, IEEE Trans. Commun. (2024). doi:10.1109/tcomm.2024.3420708.

[65] A. A. Alemi, I. Fischer, J. V. Dillon, K. Murphy, Deep variational information bottleneck, arXiv preprint arXiv:1612.00410 (2016).

[66] X. Chen, S. Gan, C. Feng, X. Wang, T. Q. Quek, Take what you need: Flexible multi-task semantic communications with channel adaptation, arXiv preprint arXiv:2502.08221 (2025).

[67] G. Zhang, Q. Hu, Z. Qin, Y. Cai, G. Yu, A unified multi-task semantic communication system with domain adaptation, in: Proc. GLOBECOM 2022-2022 IEEE Global Commun. Conf., IEEE, 2022, pp. 3971–3976. doi:10.1109/globecom48099.2022.10000850.

[68] H. Xie, Z. Qin, X. Tao, Z. Han, Toward intelligent communications: Large model empowered semantic communications, IEEE Commun. Mag. (2024). doi:10.1109/mcom.001.2300807.

[69] Z. Wang, Y. Deng, A. H. Aghvami, Task-oriented and semantics-aware communications for augmented reality, in: Proc. GLOBECOM 2023-2023 IEEE Global Commun. Conf., IEEE, 2023, pp. 2215–2220. doi:10.1109/globecom54140.2023.10437075.

[70] Y. Wang, S. Han, X. Xu, H. Liang, R. Meng, C. Dong, P. Zhang, Feature importance-aware task-oriented semantic transmission and optimization, IEEE Trans. Cogn. Commun. Netw. 10 (4) (2024) 1175–1189. doi:10.1109/tccn.2024.3375496.

[71] S. Ma, W. Qiao, Y. Wu, H. Li, G. Shi, D. Gao, Y. Shi, S. Li, N. Al-Dhahir, Task-oriented explainable semantic communications, IEEE Trans. Wireless Commun. 22 (12) (2023) 9248–9262. doi:10.1109/twc.2023.3269444.

[72] B. Du, H. Du, D. Niyato, R. Li, Task-oriented semantic communication in large multimodal models-based vehicle networks, IEEE Trans. Mobile Comput. (2025). doi:10.1109/tmc.2025.3564543.

[73] S. Sun, Z. Qin, H. Xie, X. Tao, Task-oriented explainable semantic communications based on structured scene graphs, in: Proc. GLOBECOM 2023-2023 IEEE Global Commun. Conf., IEEE, 2023, pp. 3222–3227. doi:10.1109/globecom54140.2023.10436793.

[74] M. Chen, M. Liu, C. Wang, X. Song, Z. Zhang, Y. Xie, L. Wang, Cross-modal graph semantic communication assisted by generative AI in the metaverse for 6G, Research 7 (2024) 0342. doi:10.34133/research.0342.

[75] S. Salehi, M. Erol-Kantarci, D. Niyato, LLM-enabled data transmission in end-to-end semantic communication, in: Proc. 2025 IEEE Symp. Comput. Commun. (ISCC), 2025, pp. 1–6. doi:10.1109/ISCC65549.2025.11326061.

[76] J. Choi, J. Park, E. Grassucci, D. Comminiello, Semantic communication challenges: Understanding dos and avoiding don'ts, in: Proc. 2024 IEEE 99th Veh. Technol. Conf. (VTC2024-Spring), IEEE, 2024, pp. 1–5. doi:10.1109/vtc2024-spring62846.2024.10683415.

[77] S. Liu, N. Li, Y. Deng, T. Q. Quek, Goal-oriented semantic communication for wireless visual question answering, IEEE J. Sel. Areas Commun. 43 (12) (2025) 4247–4261, arXiv:2411.02452. doi:10.1109/jsac.2025.3623184.

[78] Q. Zhou, R. Li, Z. Zhao, C. Peng, H. Zhang, Semantic communication with adaptive universal transformer, IEEE Wireless Commun. Lett. 11 (3) (2021) 453–457. doi:10.1109/lwc.2021.3132067.

[79] W. Zhang, H. Zhang, H. Ma, H. Shao, N. Wang, V. C. Leung, Predictive and adaptive deep coding for wireless image transmission in semantic communication, IEEE Trans. Wireless Commun. 22 (8) (2023) 5486–5501. doi:10.1109/twc.2023.3234408.

[80] J. Park, S. W. Yoon, Transmit what you need: task-adaptive semantic communications for visual information, IEEE J. Sel. Areas Commun. 43 (12) (2025) 4182–4197, arXiv:2412.13646. doi:10.1109/jsac.2025.3623159.

[81] Y. Blau, T. Michaeli, Rethinking lossy compression: The rate-distortion-perception tradeoff, in: Proc. Int. Conf. Mach. Learn., PMLR, 2019, pp. 675–685.

[82] W. Wang, C. Wang, Z. Zhang, W. Xu, P. Zhang, F. Jiang, J. Xu, Y. Cai, M. Pang, L. Xu, T. Q. S. Quek, Resilient image semantic communication based on rate-optimized information bottleneck theory, IEEE Trans. Netw. Sci. Eng. 13 (2026) 3127–3143. doi:10.1109/tnse.2025.3626331.

[83] M. Sun, D. Wu, P. Zhang, R. Wang, Disentangled information bottleneck guided multidevice cooperative task-oriented semantic communication, IEEE Trans. Commun. 74 (2026) 6259–6274. doi:10.1109/tcomm.2026.3675469.

[84] J. He, Y. Deng, S. Wu, B. Cao, S. Meng, W. Zhang, Q. Zhang, Task-adaptive semantic communication with feedback: A conditional rate-distortion approach, IEEE Trans. Cogn. Commun. Netw. 12 (2026) 7157–7171. doi:10.1109/tccn.2026.3682513.

[85] Z. Goldfeld, Y. Polyanskiy, The information bottleneck problem and its applications in machine learning, IEEE J. Sel. Areas Inf. Theory 1 (1) (2020) 19–38. doi:10.1109/jsait.2020.2991561.

[86] Y. Xiao, X. Zhang, Y. Li, G. Shi, T. Başar, Rate-distortion theory for strategic semantic communication, in: Proc. 2022 IEEE Inf. Theory Workshop (ITW), IEEE, 2022, pp. 279–284. doi:10.1109/itw54588.2022.9965825.

[87] L. Yan, Z. Qin, R. Zhang, Y. Li, G. Y. Li, Resource allocation for text semantic communications, IEEE Wireless Commun. Lett. 11 (7) (2022) 1394–1398. doi:10.1109/lwc.2022.3170849.

[88] T. Han, Q. Yang, Z. Shi, S. He, Z. Zhang, Semantic-preserved communication system for highly efficient speech transmission, IEEE J. Sel. Areas Commun. 41 (1) (2022) 245–259. doi:10.1109/jsac.2022.3221952.

[89] Z. Weng, Z. Qin, Semantic communication systems for speech transmission, IEEE J. Sel. Areas Commun. 39 (8) (2021) 2434–2444. doi:10.1109/jsac.2021.3087240.

[90] A. W. Rix, J. G. Beerends, M. P. Hollier, A. P. Hekstra, Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs, in: Proc. 2001 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Vol. 2, IEEE, 2001, pp. 749–752. doi:10.1109/icassp.2001.941023.

[91] F. Ribeiro, D. Florêncio, C. Zhang, M. Seltzer, Crowdmos: An approach for crowdsourcing mean opinion score studies, in: Proc. 2011 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), IEEE, 2011, pp. 2416–2419. doi:10.1109/icassp.2011.5946971.

[92] W. Wu, Y. Yang, Y. Deng, A. H. Aghvami, Goal-oriented semantic communications for robotic waypoint transmission: The value and age of information approach, IEEE Trans. Wireless Commun. (2024). doi:10.1109/twc.2024.3424493.
[93] X. Song, F. Zhou, R. Ding, Z. Qu, Y. Li, Q. Wu, N. Al-Dhahir, UAV cognitive semantic communications enabled by knowledge graph for robust object detection, IEEE Trans. Commun. (2025). doi:10.1109/tcomm.2025.3538850.
[94] A. Maatouk, M. Assaad, A. Ephremides, The age of incorrect information: An enabler of semantics-empowered communication, IEEE Trans. Wireless Commun. 22 (4) (2022) 2621–2635. doi:10.1109/twc.2022.3213227.
[95] L. Yan, Z. Qin, R. Zhang, Y. Li, G. Y. Li, QoE-aware resource allocation for semantic communication networks, in: Proc. GLOBECOM 2022-2022 IEEE Global Commun. Conf., IEEE, 2022, pp. 3272–3277. doi:10.1109/globecom48099.2022.10001594.
[96] L. Xia, Y. Sun, D. Niyato, X. Li, M. A. Imran, Joint user association and bandwidth allocation in semantic communication networks, IEEE Trans. Veh. Technol. 73 (2) (2023) 2699–2711. doi:10.1109/tvt.2023.3318510.
[97] D. Huang, F. Gao, X. Tao, Q. Du, J. Lu, Toward semantic communications: Deep learning-based image semantic coding, IEEE J. Sel. Areas Commun. 41 (1) (2022) 55–71. doi:10.1109/jsac.2022.3221999.
[98] L. Qiao, M. B. Mashhadi, Z. Gao, C. H. Foh, P. Xiao, M. Bennis, Latency-aware generative semantic communications with pre-trained diffusion models, IEEE Wireless Commun. Lett. (2024). doi:10.1109/lwc.2024.3429295.
[99] Y. Feng, H. Shen, Z. Shan, Q. Yang, X. Shi, Semantic communication for edge intelligence enabled autonomous driving system, IEEE Netw. 39 (2) (2025) 149–157, arXiv:2406.10606. doi:10.1109/mnet.2024.3468328.
[100] J. Lv, H. Tong, Q. Pan, Z. Zhang, X. He, T. Luo, C. Yin, Importance-aware image segmentation-based semantic communication for autonomous driving, arXiv preprint arXiv:2401.10153 (2024).
[101] C. Liu, C. Guo, Y. Yang, N. Jiang, Adaptable semantic compression and resource allocation for task-oriented communications, IEEE Trans. Cogn. Commun. Netw. (2023). doi:10.1109/tccn.2023.3346481.
[102] S. S. Hassan, L. X. Nguyen, Y. K. Tun, Z. Han, C. S. Hong, Semantic enabled 6g leo satellite communication for earth observation: A resource-constrained network optimization, in: Proc. GLOBECOM 2024-2024 IEEE Global Commun. Conf., IEEE, 2024, pp. 4472–4478. doi:10.1109/globecom52923.2024.10901032.
[103] W. Chen, C. Ju, T. Yuan, Y. Zhan, M. Zhang, D. Wang, Free space optical semantic communication for satellite remote sensing image transmission, IEEE Trans. Commun. (2025). doi:10.1109/tcomm.2025.3562356.
[104] Z. Xu, J. Jiao, G. Yuan, S. Jiang, W. Wang, Y. Wang, Q. Zhang, Task-oriented semantic coding and utility optimal transmission in satellite-integrated internet, IEEE Trans. Geosci. Remote Sens. (2025). doi:10.1109/tgrs.2025.3576382.
[105] X. Song, L. Yuan, Z. Qu, F. Zhou, Q. Wu, T. Q. Quek, R. Q. Hu, Knowledge graph driven UAV cognitive semantic communication systems for efficient object detection, in: Proc. ICC 2024 - IEEE Int. Conf. Commun., IEEE, 2024, pp. 1685–1690, arXiv:2401.13995. doi:10.1109/icc51166.2024.10622874.
[106] J. Xu, H. Yao, R. Zhang, T. Mai, S. Huang, S. Guo, Federated learning powered semantic communication for UAV swarm cooperation, IEEE Wireless Commun. 31 (4) (2024) 140–146. doi:10.1109/mwc.005.2300147.
[107] Y. Xu, H. Zhou, Y. Deng, Task-oriented semantics-aware communication for wireless UAV control and command transmission, IEEE Commun. Lett. 27 (8) (2023) 2232–2236. doi:10.1109/lcomm.2023.3290109.
[108] P. Shah, A. Garg, V. Gajjar, Per-vis: Person retrieval in video surveillance using semantic description, in: Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis., 2021, pp. 41–50. doi:10.1109/wacvw52041.2021.00009.
[109] J. Shao, X. Zhang, J. Zhang, Task-oriented communication for edge video analytics, IEEE Trans. Wireless Commun. (2023). doi:10.1109/twc.2023.3314888.
[110] K. Nandakishore, R. V. S. Rithvik, M. Z. A. Khan, Efficient image transmission using semantic communication for static environment surveillance, in: Proc. 2024 IEEE Int. Conf. Adv. Netw. Telecommun. Syst. (ANTS), IEEE, 2024, pp. 1–6. doi:10.1109/ants63515.2024.10898424.
[111] A. Y. Huang, Y. Chen, D. Huang, M. Zhao, Semantic privacy-preserving for video surveillance services on the edge, in: Proc. 8th ACM/IEEE Symp. Edge Comput., 2023, pp. 300–305. doi:10.1145/3583740.3626820.
[112] Y. Yang, C. Guo, F. Liu, C. Liu, L. Sun, Q. Sun, J. Chen, Semantic communications with AI tasks, arXiv preprint arXiv:2109.14170 (2021).
[113] H. Xie, Z. Qin, A lite distributed semantic communication system for Internet of Things, IEEE J. Sel. Areas Commun. 39 (1) (2020) 142–153. doi:10.1109/jsac.2020.3036968.
[114] Q. Hu, G. Zhang, Z. Qin, Y. Cai, G. Yu, G. Y. Li, Robust semantic communications with masked VQ-VAE enabled codebook, IEEE Trans. Wireless Commun. 22 (12) (2023) 8707–8722. doi:10.1109/twc.2023.3265201.
[115] H. Chen, F. Fang, X. Wang, Semantic extraction model selection for IoT devices in edge-assisted semantic communications, IEEE Commun. Lett. (2024). doi:10.1109/lcomm.2024.3394323.
[116] S. A. Moqurrab, A. Anjum, N. Tariq, G. Srivastava, Instant_Anonymity: A lightweight semantic privacy guarantee for 5G-enabled IIoT, IEEE Trans. Ind. Inf. 19 (1) (2022) 951–959. doi:10.1109/tii.2022.3179536.
[117] A. Willner, C. Diedrich, R. B. Younes, S. Hohmann, A. Kraft, Semantic communication between components for smart factories based on oneM2M, in: Proc. 2017 22nd IEEE Int. Conf. Emerg. Technol. Factory Autom. (ETFA), IEEE, 2017, pp. 1–8. doi:10.1109/etfa.2017.8247690.
[118] Z. Zhao, Z. Yang, C. Huang, L. Wei, Q. Yang, C. Zhong, W. Xu, Z. Zhang, A joint communication and computation design for distributed RISs assisted probabilistic semantic communication in IIoT, IEEE Internet Things J. (2024). doi:10.1109/jiot.2024.3409271.
[119] M. Zhang, Y. Yang, B. Song, Z. Chen, X. Ma, D. Li, N. Li, F. Cao, Y. Zhang, H. Zhang, Industrial semantic communications for tool wear monitoring in industrial IoT networks, in: Proc. 2024 IEEE/CIC Int. Conf. Commun. China (ICCC Workshops), IEEE, 2024, pp. 248–253. doi:10.1109/icccworkshops62562.2024.10693780.
[120] Y. Yang, Z. Yang, W. Yuan, F. Liu, X. Cao, C. Huang, Z. Zhang, M. Shikh-Bahaei, E-healthcare systems: Integrated sensing, computing, and semantic communication with physical layer security, arXiv preprint arXiv:2409.20453 (2024).
[121] V. K. Bura, B. Jamalpur, DeepSemComm: a deep learning-based semantic communication framework with case studies in text and biomedical signal transmission, Eur. Phys. J. Plus 140 (6) (2025) 597. doi:10.1140/epjp/s13360-025-06520-9.

[122] M. Yang, D. Gao, J. Li, W. Xu, G. Shi, Superimposed semantic communication for iot-based real-time ecg monitoring, IEEE J. Biomed. Health Inform. 28 (7) (2024) 3819–3830. doi:10.1109/jbhi.2024.3352927.

[123] P. Li, A. Aijaz, Task-oriented connectivity for networked robotics with generative AI and semantic communications, in: Proc. IEEE INFOCOM 2025 - IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS), IEEE, 2025, pp. 1–6, arXiv:2503.06771. doi:10.1109/infocomwkshps65812.2025.11152855.

[124] D. Wu, Q. Zhao, J. Fan, J. Qi, P. Zheng, J. Hu, H2R bridge: Transferring vision-language models to few-shot intention meta-perception in human robot collaboration, J. Manuf. Syst. 80 (2025) 524–535. doi:10.1016/j.jmsy.2025.03.016.

[125] A. Suresh, E. Beck, A. Dekorsy, P. Rückert, K. Tracht, Human-integrated multi-agent exploration using semantic communication and extended reality simulation, in: Proc. 2024 10th Int. Conf. Autom., Robot. Appl. (ICARA), Athens, Greece, 2024, pp. 22–24. doi:10.1109/icara60736.2024.10553050.

[126] S. Chen, E. Spyrakos-Papastavridis, Y. Jin, Y. Deng, Goal-oriented semantic communication for robot arm reconstruction in digital twin: Feature and temporal selections, IEEE J. Sel. Areas Commun. 43 (9) (2025) 3072–3087. doi:10.1109/JSAC.2025.3574592.

[127] J. Tang, J. Nie, J. Bai, J. Xu, S. Li, Y. Zhang, Y. Yuan, UAV-assisted digital twin synchronization with tiny machine learning-based semantic communications, IEEE Internet Things J. (2024). doi:10.1109/jiot.2024.3401229.

[128] B. Li, H. Cai, L. Liu, Z. Fei, Delay-aware digital twin synchronization in mobile edge networks with semantic communications, IEEE Trans. Veh. Technol. (2025). doi:10.1109/tvt.2025.3548844.

[129] T. Kreuzer, P. Papapetrou, J. Zdravkovic, Artificial intelligence in digital twins—a systematic literature review, Data Knowl. Eng. 151 (2024) 102304. doi:10.1016/j.datak.2024.102304.

[130] Y. Huang, B. Bai, Y. Zhu, X. Qiao, X. Su, L. Yang, P. Zhang, ISCom: Interest-aware semantic communication scheme for point cloud video streaming on metaverse XR devices, IEEE J. Sel. Areas Commun. 42 (4) (2023) 1003–1021. doi:10.1109/jsac.2023.3345430.

[131] W. Yang, Z. Xiong, T. Q. Quek, X. Shen, Streamlined transmission: A semantic-aware XR deployment framework enhanced by generative AI, IEEE Netw. 38 (6) (2024) 29–38. doi:10.1109/mnet.2024.3414398.

[132] G. Liu, H. Du, D. Niyato, J. Kang, Z. Xiong, B. H. Soong, Vision-based semantic communications for metaverse services: A contest theoretic approach, in: Proc. GLOBECOM 2023-2023 IEEE Global Commun. Conf., IEEE, 2023, pp. 2426–2432. doi:10.1109/globecom54140.2023.10437058.

[133] Y. Zhang, Y. Sun, L. Guo, W. Chen, B. Ai, D. Gunduz, Multimodal llm integrated semantic communications for 6g immersive experiences, IEEE Commun. Mag. 64 (5) (2026) 68–75, arXiv:2507.04621. doi:10.1109/mcom.001.2500420.

[134] Y. Li, S. Jiang, B. Liu, B. Wang, L. Wang, M. Rao, Semantic communication-enabled cloud-edge-end-collaborative metaverse services architecture, arXiv preprint arXiv:2506.10001 (2025).

[135] C. Liu, C. Guo, Y. Yang, W. Ni, Y. Zhou, L. Li, T. Q. Quek, Explainable semantic communication for text tasks, IEEE Internet Things J. (2024). doi:10.1109/jiot.2024.3450902.

[136] J. Shi, T.-T. Chan, H. Pan, T.-M. Lok, Reconfigurable intelligent surface assisted semantic communication systems, in: Proc. 2023 IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC), IEEE, 2023, pp. 1–6. doi:10.1109/icspcc59353.2023.10400366.

[137] W. Han, A. Zhang, C. Feng, R. Chen, S. Zhang, S. Guo, Few-shot adaptive learning for robust task-oriented semantic communication, in: Proc. AAAI 2025 Workshop Artif. Intell. Wireless Commun. Netw. (AI4WCN), 2025.

[138] S. Chen, E. Spyrakos-Papastavridis, Y. Jin, Y. Deng, Goal-oriented semantic communication for robotic arm reconstruction in digital twin, in: Proc. GLOBECOM 2024-2024 IEEE Global Commun. Conf., IEEE, IEEE, 2024, pp. 4830–4835. doi:10.1109/globecom52923.2024.10901815.

[139] S. D. Okegbile, H. Gao, J. Cai, A novel secure split federated semantic learning framework and its optimization for digital twin network evolution, IEEE Trans. Mobile Comput. 25 (1) (2026) 1302–1319. doi:10.1109/tmc.2025.3599838.

[140] Z. Lv, S. Xie, Artificial intelligence in the digital twins: State of the art, challenges, and future research topics: [version 2; peer review: 2 approved], Digit. Twin 1 (1) (2024) 12. doi:10.12688/digitaltwin.17524.2.

[141] S. Ma, K. A. Flanigan, M. Bergés, State-of-the-art review: The use of digital twins to support artificial intelligence-guided predictive maintenance, arXiv preprint arXiv:2406.13117 (2024).

[142] N. Farsad, M. Rao, A. Goldsmith, Deep learning for joint source-channel coding of text, in: Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), IEEE, 2018, pp. 2326–2330. doi:10.1109/icassp.2018.8461983.

[143] A. Défossez, J. Copet, G. Synnaeve, Y. Adi, High fidelity neural audio compression, Trans. Mach. Learn. Res. (2023).

[144] J. Zhang, S. Li, Y. Sun, K. Cai, H. Chen, X. Xu, S. Cui, Cooperative semantic knowledge base update for semantic communication networks, IEEE Trans. Commun. 74 (2026) 1391–1405. doi:10.1109/tcomm.2025.3636085.

[145] W. Yang, Z. Xiong, S. Mao, T. Q. Quek, P. Zhang, M. Debbah, R. Tafazolli, Rethinking generative semantic communication for multi-user systems with large language models, IEEE Wireless Commun. (2025). doi:10.1109/mwc.002.2400332.

[146] N. Li, Y. Deng, D. Niyato, Goal-oriented semantic communication for wireless video transmission via generative AI, IEEE Trans. Wireless Commun. 25 (2026) 10841–10854. doi:10.1109/twc.2026.3654401.

[147] S. Ribouh, O. Saleem, Large language model-based semantic communication system for image transmission, arXiv preprint arXiv:2501.12988 (2025).

[148] L. Yao, Y. Wei, Y. Zhou, Z. Li, N. Chi, K.-K. Wong, You only transmit once: Unified generation and comprehension for efficient semantic communication, IEEE J. Sel. Areas Commun. 44 (2026) 3349–3364. doi:10.1109/jsac.2026.3655643.

[149] J. Ding, P. Jiang, C.-K. Wen, X. Li, S. Jin, Foundation model-aided channel-adaptive video semantic communication and prototype validation, IEEE Trans. Wireless Commun. 25 (2026) 2892–2906. doi:10.1109/twc.2025.3600221.

[150] H. Zhang, B. Di, H. Zhang, D. Niyato, L. Song, Feature-foundation model evolution for low-latency semantic communication, IEEE Trans. Wireless Commun. 25 (2026) 11915–11929. doi:10.1109/TWC.2026.3659294.

[151] Y. Sun, Y. Liu, S. Guo, X. Qiu, J. Chen, J. Hao, D. Niyato, Edge large AI model agent-empowered cognitive multimodal semantic communication, IEEE Trans. Mobile Comput. 25 (1) (2026). doi:10.1109/TMC.2025.3590723.

[152] A. Islam, K. Chang, Beyond bit fidelity: Dual-mode domain-aware knowledge-enhanced semantic text communication framework, IEEE Trans. Netw. Sci. Eng. 13 (2026) 7964–7986. doi:10.1109/tnse.2026.3676381.

[153] C. Liu, C. Guo, Y. Yang, M. Chen, T. Q. Quek, Lightweight task-oriented semantic communication empowered by large-scale AI models, IEEE Trans. Veh. Technol. (2025). doi:10.1109/tvt.2025.3563635.
[154] J. Peng, H. Xing, Y. Li, L. Feng, L. Xu, X. Lei, Task-oriented multi-user semantic communication with lightweight semantic encoder and fast training for resource-constrained terminal devices, IEEE Wireless Commun. Lett. (2024). doi:10.1109/lwc.2024.3417028.
[155] S. Guo, A. Zhang, Y. Wang, C. Feng, T. Q. Quek, Semantic-enabled 6g communication: A task-oriented and privacy-preserving perspective, IEEE Netw. (2025). doi:10.1109/mnet.2025.3547760.
[156] C. K. Thomas, W. Saad, Neuro-symbolic artificial intelligence (AI) for intent based semantic communication, in: Proc. GLOBECOM 2022-2022 IEEE Global Commun. Conf., IEEE, 2022, pp. 2698–2703. doi:10.1109/globecom48099.2022.10001097.
[157] C. Aceta, I. Fernández, A. Soroa, KIDE4I: A generic semantics-based task-oriented dialogue system for human-machine interaction in Industry 5.0, Appl. Sci. 12 (3) (2022) 1192. doi:10.3390/app12031192.
[158] E. Grassucci, F. Venturini, D. Comminiello, Generative semantic communication: Diffusion models beyond bit recovery, in: Proc. IEEE Global Commun. Conf. (GLOBECOM), IEEE, 2023, pp. 3603–3608.
[159] M. Yang, B. Liu, B. Wang, H.-S. Kim, Diffusion-aided joint source channel coding for high realism wireless image transmission, IEEE Trans. Mach. Learn. Commun. Netw. 3 (2025) 1227–1243. doi:10.1109/TMLCN.2025.3628535.
[160] T. Wu, Z. Chen, D. He, L. Qian, Y. Xu, M. Tao, W. Zhang, CDDM: Channel denoising diffusion models for wireless semantic communications, IEEE Trans. Wireless Commun. 23 (9) (2024) 11168–11183. doi:10.1109/TWC.2024.3379244.
[161] Z. Fang, M. Sun, S. Wang, X. Xu, H. Gao, J. Huang, S. Han, P. Zhang, Cross-modal generative semantic communications powered by semantic knowledge base, IEEE Trans. Netw. Sci. Eng. 13 (2026) 5568–5585. doi:10.1109/tnse.2025.3650269.
[162] Y. Ma, H. Yin, L. Qiao, S. Sun, Z. Gao, Y. Xu, W. Zhang, LGVSC: A large-model-driven generative video semantic communication framework, IEEE Trans. Veh. Technol. (2026) 1–6doi:10.1109/tvt.2026.3705309.
[163] H. Zhang, H. Wang, Y. Li, K. Long, A. Nallanathan, DRL-driven dynamic resource allocation for task-oriented semantic communication, IEEE Trans. Commun. 71 (7) (2023) 3992–4004. doi:10.1109/tcomm.2023.3274145.
[164] Z. Xiang, Y. Li, F. Yu, Y. Wang, Y. Li, Z. Rao, Task-oriented semantic communication via federated learning, in: Proc. 2025 IEEE Int. Perform., Comput., Commun. Conf. (IPCCC), IEEE, 2025, pp. 1–7. doi:10.1109/ipccc66453.2025.11304693.
[165] A. H. Razlighi, C. Bockelmann, A. Dekorsy, Semantic communication for cooperative multi-task processing over wireless networks, IEEE Wireless Commun. Lett. (2024). doi:10.1109/LWC.2024.3451139.
[166] A. H. Razlighi, M. H. Tillmann, E. Beck, C. Bockelmann, A. Dekorsy, Cooperative and collaborative multi-task semantic communication for distributed sources, in: Proc. ICC 2025-IEEE Int. Conf. Commun., IEEE, 2025, pp. 3966–3971. doi:10.1109/icc52391.2025.11161271.
[167] G. He, Exploring task-oriented communication in multi-agent system: A deep reinforcement learning approach, arXiv preprint arXiv:2208.10165 (2022).
[168] Y. Cang, M. Chen, Z. Yang, Y. Hu, Y. Wang, Z. Zhang, K.-K. Wong, Resource allocation for semantic-aware mobile edge computing systems, in: Proc. 2023 IEEE Globecom Workshops (GC Wkshps), IEEE, 2023, pp. 1585–1590. doi:10.1109/gcwkshps58843.2023.10464965.
[169] J. Guo, H. Chen, B. Song, Y. Chi, C. Yuen, F. R. Yu, G. Y. Li, D. Niyato, Distributed task-oriented communication networks with multimodal semantic relay and edge intelligence, IEEE Commun. Mag. (2024). doi:10.1109/mcom.001.2300155.
[170] J. Guo, C. Yang, Learning precoding for semantic communications, in: Proc. 2022 IEEE Int. Conf. Commun. Workshops (ICC Workshops), IEEE, 2022, pp. 163–168. doi:10.1109/iccworkshops53468.2022.9814464.
[171] Y. Li, Z. Jin, T. Song, X. Song, J. Hu, Robust task-oriented communication with semantic-aware masking and discrete codebook, in: Proc. 2025 IEEE Wireless Commun. Netw. Conf. (WCNC), IEEE, 2025, pp. 1–6. doi:10.1109/wcnc61545.2025.10978541.
[172] C. Xu, Y. Ma, R. Tafazolli, J. Wang, Data-importance-aware power allocation for adaptive real-time communication in computer vision applications, IEEE J. Sel. Areas Commun. 43 (12) (2025) 4015–4026, arXiv:2504.08922. doi:10.1109/jsac.2025.3623165.
[173] N. Gholami, N. Moghim, B. S. Ghahfarokhi, P. Salavati, C. K. Thomas, S. Shetty, T. Rahmati, Flexible semantic-aware resource allocation: Serving more users through similarity range constraints, arXiv preprint arXiv:2504.20939 (2025).
[174] S. Vaidanis, P. A. Stavrou, M. Kountouris, Optimization for semantic-aware resource allocation under CPT-based utilities, in: Proc. 2025 IEEE 26th Int. Workshop Signal Process. Artif. Intell. Wireless Commun. (SPAWC), IEEE, 2025, pp. 1–5, arXiv:2506.04952. doi:10.1109/spawc66079.2025.11143243.
[175] T. T. Nguyen, T.-D. Le, V. N. Ha, H.-f. Chou, G. Eappen, D.-D. Tran, H. Nguyen-Kha, P. Thiruvasagam, L. M. Garces-Socarras, J. L. Gonzalez-Rios, et al., A semantic-loss function modeling framework with task-oriented machine learning perspectives, in: Proc. 2025 IEEE Int. Conf. Mach. Learn. Commun. Netw. (ICMLCN), IEEE, 2025, pp. 1–6, arXiv:2503.09903. doi:10.1109/icmlcn64995.2025.11140499.
[176] J. Ma, Q. Li, R. Liu, A. Pandharipande, X. Ge, Enhanced semantic information transfer on RIS-assisted communication systems, IEEE Wireless Commun. Lett. 13 (8) (2024) 2225–2229. doi:10.1109/lwc.2024.3408420.
[177] S. Guo, J. Zhou, Z. Li, P. Zhang, S. Zhang, Reconfigurable intelligent surfaces-assisted task-oriented communications for AI-driven vertical applications, in: Proc. 2024 22nd Int. Symp. Model. Optim. Mobile, Ad Hoc, Wireless Netw. (WiOpt), IEEE, 2024, pp. 99–104.
[178] P. Jiang, C.-K. Wen, S. Jin, G. Y. Li, RIS-enhanced semantic communications adaptive to user requirements, IEEE Trans. Commun. (2024). doi:10.1109/tcomm.2024.3369697.
[179] G. Huang, J. An, Z. Yang, L. Gan, M. Bennis, M. Debbah, Stacked intelligent metasurfaces for task-oriented semantic communications, IEEE Wireless Commun. Lett. (2024). doi:10.1109/lwc.2024.3499970.
[180] X. Hu, Y. Tian, Q. Li, Y. H. Kho, X. Wang, B. Xiao, Z. Yang, W. Li, A novel RIS-aided optimization strategy for semantic communication system, IEEE Wireless Commun. Lett. 13 (6) (2024) 1655–1659. doi:10.1109/lwc.2024.3385492.
[181] S. Mao, C. Yuen, L. Liu, M. Xiao, S. Yu, N. Zhang, RIS-enhanced semantic-aware sensing, communication, computation, and control for Internet of Things, IEEE Trans. Wireless Commun. 25 (2026) 2231–2246. doi:10.1109/twc.2025.3595550.
[182] M. Liu, J. An, C. Huang, C. Yuen, Over-the-air ode-inspired neural network for dual task-oriented semantic communications, IEEE Trans. Cogn. Commun. Netw. (2025). doi:10.1109/tccn.2025.3569604.
[183] Y. Chen, Q. Yang, Z. Shi, J. Chen, The model inversion eavesdropping attack in semantic communication systems, in: Proc. IEEE Global Commun. Conf. (GLOBECOM), 2023. doi:10.1109/GLOBECOM54140.2023.10436996.

[184] Y. Wang, S. Guo, Y. Deng, H. Zhang, Y. Fang, Privacy-preserving task-oriented semantic communications against model inversion attacks, IEEE Trans. Wireless Commun. (2024). doi:10.1109/twc.2024.3369170.
[185] O. Erak, O. Alhussein, W. Tong, Contrastive learning and adversarial disentanglement for privacy-preserving task-oriented semantic communications, arXiv preprint arXiv:2410.22784 (2024).
[186] E. Erdemir, P. L. Dragotti, D. Gündüz, Privacy-aware communication over a wiretap channel with generative networks, in: Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), IEEE, 2022, pp. 2989–2993. doi:10.1109/icassp43922.2022.9747068.
[187] J. Xu, X. Wang, L. Chen, L. Lin, J. Hu, S. A. Khowaja, K. Dev, Fortifying secure semantic communication: A next-generation defense framework against model inversion and adversarial threats, IEEE Wireless Commun. 32 (5) (2025) 64–71. doi:10.1109/MWC.2500044.
[188] J. Peng, H. Xing, L. Xu, S. Luo, P. Dai, L. Feng, J. Song, B. Zhao, Z. Xiao, Adversarial reinforcement learning based data poisoning attacks defense for task-oriented multi-user semantic communication, IEEE Trans. Mobile Comput. 23 (12) (2024) 14834–14851. doi:10.1109/TMC.2024.3447087.
[189] X. Yang, Y. Lai, G. Li, J. Wu, K. Zhou, J. Li, M. Chen, Toward polymorphic backdoor against semantic communication via intensity-based poisoning, IEEE Trans. Inf. Forensics Security (2026). doi:10.1109/TIFS.2026.3689282.
[190] Z. Wei, Y. Jiang, J. Huang, F. Zhong, S. Gyawali, Detecting backdoor attacks via similarity in semantic communication systems, in: Lecture Notes in Networks and Systems, Springer, 2025, pp. 315–329, arXiv:2502.03721. doi:10.1007/978-3-032-09694-4_25.
[191] M. Sadeghi, E. G. Larsson, Adversarial attacks on deep-learning based radio signal classification, IEEE Wireless Commun. Lett. 8 (1) (2019) 213–216. doi:10.1109/LWC.2018.2867459.
[192] B. Flowers, R. M. Buehrer, W. C. Headley, Evaluating adversarial evasion attacks in the context of wireless communications, IEEE Trans. Inf. Forensics Security 15 (2020) 1102–1113. doi:10.1109/TIFS.2019.2934069.
[193] B. Kim, Y. E. Sagduyu, K. Davaslioglu, T. Erpek, S. Ulukus, Channel-aware adversarial attacks against deep learning-based wireless signal classifiers, IEEE Trans. Wireless Commun. 21 (6) (2022) 3868–3880. doi:10.1109/TWC.2021.3124855.
[194] K. Wei, R. Xie, W. Xu, Z. Lu, H. Xiao, Robust semantic communication via adversarial training, IEEE Trans. Veh. Technol. 74 (12) (2025) 19849–19853. doi:10.1109/TVT.2025.3585536.
[195] D. T. M. Hoang, L. B. Le, Backdoor attacks on large language model based semantic communication systems, in: Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops), 2025, pp. 518–523. doi:10.1109/ICCWorkshops67674.2025.11162215.
[196] M. Fredrikson, S. Jha, T. Ristenpart, Model inversion attacks that exploit confidence information and basic countermeasures, in: Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur., 2015, pp. 1322–1333. doi:10.1145/2810103.2813677.
[197] R. Shokri, M. Stronati, C. Song, V. Shmatikov, Membership inference attacks against machine learning models, in: Proc. 2017 IEEE Symp. Secur. Priv. (SP), IEEE, 2017, pp. 3–18. doi:10.1109/SP.2017.41.
[198] F. Tramèr, F. Zhang, A. Juels, M. K. Reiter, T. Ristenpart, Stealing machine learning models via prediction APIs, in: Proc. 25th USENIX Secur. Symp. (USENIX Security 16), 2016, pp. 601–618.
[199] X. Yang, Y. Lai, G. Li, J. Wu, K. Zhou, M. Chen, SemanAegis: Toward credential-aware semantic communication against knowledge leakage threats, IEEE Trans. Mobile Comput. 25 (4) (2026) 5032–5049. doi:10.1109/tmc.2025.3627473.
[200] L. Zhu, Z. Liu, S. Han, Deep leakage from gradients, in: Adv. Neural Inf. Process. Syst., Vol. 32, 2019.
[201] I. Shumailov, Y. Zhao, D. Bates, N. Papernot, R. Mullins, R. Anderson, Sponge examples: Energy-latency attacks on neural networks, in: Proc. 2021 IEEE Eur. Symp. Secur. Priv. (EuroS&P), IEEE, 2021, pp. 212–231. doi:10.1109/EuroSP51992.2021.00024.
[202] L. Hu, H. Wang, T. Xiang, Z. Chen, J. He, S. Guo, Vector-guided embezzler: A reconstruction attack against retrieval-augmented generation in agentic AI communication, IEEE Trans. Cogn. Commun. Netw. 12 (2026) 8303–8316. doi:10.1109/tccn.2026.3694035.
[203] J. He, X. Luo, J. Kang, Z. Xiong, X. Li, C. Chen, D. Niyato, Mixture of experts-based semantic communication for secure Artificial Intelligence of Things, IEEE Trans. Veh. Technol. 75 (2) (2026) 3324–3329. doi:10.1109/tvt.2025.3599157.
[204] N. Fayaz, R. K. Mallik, B. Lall, Semantic noise decomposition and saliency-weighted mutual information in task-oriented communication, IEEE Commun. Lett. 30 (2026) 2148–2152. doi:10.1109/lcomm.2026.3698832.
[205] R. L. Dobrushin, B. S. Tsybakov, Information transmission with additional noise, IEEE Trans. Inf. Theory 8 (5) (1962) 293–304. doi:10.1109/tit.1962.1057738.
[206] H. S. Witsenhausen, Indirect rate distortion problems, IEEE Trans. Inf. Theory 26 (5) (1980) 518–521. doi:10.1109/tit.1980.1056251.
[207] T. A. Courtade, T. Weissman, Multiterminal source coding under logarithmic loss, IEEE Trans. Inf. Theory 60 (1) (2014) 740–761.
[208] P. A. Stavrou, M. Kountouris, The role of fidelity in goal-oriented semantic communication: A rate distortion approach, IEEE Trans. Commun. 71 (7) (2023) 3881–3896. doi:10.1109/TCOMM.2023.3274122.
[209] J. Liu, S. Shao, W. Zhang, H. V. Poor, An indirect rate-distortion characterization for semantic sources: General model and the case of gaussian observation, IEEE Trans. Commun. 70 (9) (2022) 5946–5959. doi:10.1109/tcomm.2022.3194978.
[210] H. Li, J. Shao, H. He, S. Song, J. Zhang, K. B. Letaief, Tackling distribution shifts in task-oriented communication with information bottleneck, IEEE J. Sel. Areas Commun. (2025). doi:10.1109/jsac.2025.3559116.
[211] M. Kalfa, M. Gok, A. Atalik, B. Tegin, T. M. Duman, O. Arikan, Towards goal-oriented semantic signal processing: Applications and future challenges, Digit. Signal Process. 119 (2021) 103134. doi:10.1016/j.dsp.2021.103134.
[212] V. Monga, Y. Li, Y. C. Eldar, Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing, IEEE Signal Process. Mag. 38 (2) (2021) 18–44. doi:10.1109/MSP.2020.3016905.

Supplementary Material for "When Meaning Meets Purpose: A Survey on Semantic and Task-Oriented Communications"

Quang Tuan Do[a], Tung Son Do[a], Thanh Phung Truong[a], Dongwook Won[a], Nhu-Ngoc Dao[c], Wonjong Noh[b], Sungrae Cho[a]

[a]*Department of Computer Science and Engineering, Chung-Ang University, Seoul, 06974, South Korea*
[b]*School of Software, Hallym University, Chuncheon, 24252, South Korea*
[c]*Department of Computer Science and Engineering, Sejong University, Seoul, 05006, South Korea*

This document collects supplementary tables referenced in the main manuscript (*When Meaning Meets Purpose: A Survey on Semantic and Task-Oriented Communications*). Supplementary Tables S1–S4 are extended reference catalogs whose synthesis conclusions are presented in the main text. Table and citation numbering in this document is independent of the main manuscript.

Table S1: Comparative Summary of Foundational Communication Theories

Feature	**Classical Information Theory**	**Semantic Information Theory**	**Strongly Semantic Information Theory**	**Semantic Communication Theory**	**Goal-Oriented Communication Theory**
Primary Level (Weaver)	Level A (Technical)	Level B (Semantic)	Level B (Semantic)	Level B (Semantic)	Level C (Effectiveness)
Focus	Accurate symbol transmission	Quantifying semantic content implied by language elements	Quantifying truthful semantic content as discrepancy from the actual situation	Conveying meaningful content via a system model with knowledge and inference	Achieving specific goals among participants through communication
Truthfulness	Not considered	Not inherently considered in early forms	Central to information quantification	Implicitly handled by alignment with world model and knowledge	Relevant when it impacts goal achievement
Context/Knowledge	Abstracted away	Based on a defined language model	Truth relative to "actual situation"	Essential (shared KBs, inference) for encoding, decoding, and capacity	Can be established via learning or intermediaries; common language not always pre-assumed
Goal-Orientation	No	No, focused on meaning quantification	No, focused on truthful meaning quantification	Indirect, by ensuring intended meaning is conveyed for potential action	Explicitly central; communication is a means to an end
Key Advancements	Mathematical theory of communication	First attempts to formalize semantic information	Introduction of truth values into semantic measures	System-level model for semantic exchange, concept of semantic noise/capacity	Communication framed by participant goals, potential for understanding without shared language
Noted Limitations	Ignores meaning, inefficient for intelligent systems	Does not distinguish from CIT in purpose, neglects truth initially	Limited to specific logical propositions, broader rigor required	Conceptual, ideal conditions, reliance on Shannon entropy hindering deeper meaning focus	Abstract early models, trial and error paradigm

Table S2: Cross-mechanism synthesis: recurring semantic processing mechanisms across architectures and domains.

Mechanism	Representative Systems	Application Domains	Why It Recurs
IB-based extraction	Deep variational information bottleneck (VIB) [1]; IB-guided semantic encoder [2]; IB formulation for wireless SC [3]	Text, image, multi-modal, edge inference	Provides a principled, task-agnostic objective for discarding task-irrelevant information while bounding the rate-relevance trade-off.
Saliency ranking / prioritization	Personalized saliency encoder [4]; VQA object ranking [5]; skeleton joint prioritization [6]	Image classification, VQA, AR/VR, autonomous driving	Task performance is dominated by a small subset of semantic elements; ranking enables graceful degradation under bandwidth constraints.
Adaptive transmission	Channel-adaptive rate control [7]; SNR-aware semantic coding [8]	All wireless domains (vehicular, IoT, mobile edge)	Channel conditions and task urgency fluctuate in real time; static compression forgoes significant efficiency or reliability.
Feedback coupling	Receiver-driven semantic requests [9]; query-driven task communication [5]; KB-anchored semantic-model update policies [10]	Multi-agent systems, VQA, collaborative perception	Closed-loop feedback aligns the transmitter's semantic selection with the receiver's actual information needs, reducing unnecessary transmissions.
Robustness control	Adversarial training for semantic noise [11]; rate-distortion-perception trade-off [12]; channel-robust JSCC designs	Safety-critical (medical, vehicular), contested spectrum	Aggressive semantic compression amplifies sensitivity to channel errors and distribution shift; robustness mechanisms restore reliability margins.

Table S3: Summary of Performance Evaluation Metrics for Semantic Communication Systems

Metric Category	**Metric Name**	**Definition**	**Relevant Tasks**
Level B: Semantic Fidelity and Quality			
Image/Video	PSNR / SSIM / LPIPS / FID	PSNR measures pixel-level error; SSIM assesses perceptual similarity in luminance, contrast, and structure; LPIPS and FID are learned perceptual metrics capturing feature-space and distributional fidelity (Section 4 of the main text).	Image/Video Reconstruction
Text	BLEU / Semantic Similarity	BLEU counts n-gram overlap. Semantic Similarity measures cosine similarity between sentence embeddings.	Text Transmission, Machine Translation
Speech	WER / SDR / PESQ / MOS	Metrics for transcription error (WER), signal distortion (SDR), and perceptual quality (PESQ, MOS).	Speech-to-Text, Speech-to-Speech
Level C: Task Effectiveness and Goal Achievement (ToSC)			
Value, Timeliness, and Utility	VoI (Value of Information)	Quantifies the task-specific usefulness of information, often as the expected improvement in a utility function.	Control Systems, Decision Making
	AoI (Age of Information)	Measures information freshness; the time elapsed since the generation of the most recently received update.	Real-time Status Monitoring
	AoII (Age of Incorrect Information)	Applies a penalty that grows with the time elapsed since the received information became incorrect.	Semantics-aware Monitoring
Task-Specific Success	Classification Accuracy	Proportion of correctly classified samples. Often optimized via a loss function like cross-entropy.	Classification, Recognition
	mAP / AP	Mean Average Precision / Average Precision. Summarizes the precision-recall curve, typically based on an IoU threshold.	Object Detection
	MPJPE / P2Point Error	Measures Euclidean distance for articulated joints (MPJPE) or geometric distortion for 3D point clouds (P2Point).	3D Pose/Scene Reconstruction
User-Centric & System-Wide	QoE (Quality-of-Experience)	A user-centric metric of subjective satisfaction, often a composite of task effectiveness and efficiency.	Multi-task/Multi-user Networks
	STM (System Throughput in Message)	The aggregate message rate across all users in a network, derived from a bit-rate-to-message-rate (B2M) transformation.	Network-level Resource Management
Communication Efficiency and Resource Utilization			
	Bandwidth / Data Rate	Measures data volume, e.g., as bits per pixel (bpp) or a compression ratio.	General Communication Efficiency
	S-R / S-SE	Semantic Transmission Rate (suts/s) / Semantic Spectral Efficiency (suts/s/Hz). Measures effective meaning transmission rate.	Semantic-aware Networks
	End-to-end Latency	Total time from source generation to task execution, including processing and transmission delays.	Real-time Interactive Systems

Table S4: Comparative Framework of Research Trends

Research Trends	Communication Gain Approach	Computational Load	Typical Task Types	Key Limitations
LLM/LMM Integration	Massive *cross-modal* data reduction by transforming data modalities (e.g., image to text), a distinct mechanism from the within-modality codec compression that is rated moderate for this family in the main text's comparison of AI model families.	High, often requiring powerful cloud/server backends.	Flexible and complex tasks: image transmission, VQA, task decomposition.	Reliance on computationally intensive foundation models; potential scalability challenges.
Intent-based Systems	Transmits the underlying purpose ("why") of the communication, reducing data to its most concise representation.	High, as it requires end-nodes to be intelligent agents capable of reasoning.	Human-machine dialogue; tasks requiring generalization and causal reasoning.	Requires a shift from passive transceivers to intelligent agents; can be challenged by data scarcity in specialized domains.
Multi-user Systems	Efficient resource management to mitigate inter-user interference and effectively fuse complementary data from multiple sources.	High, due to joint optimization of transceivers and complex data fusion at the receiver.	Single-modal (image retrieval, machine translation) and multimodal (VQA) tasks.	Managing inter-user interference; effective fusion of diverse data types; resource constraints on end devices.
Cooperative Agents	Enables efficient sharing of distributed observations by exploiting correlations between agent tasks.	High, involving complex architectures such as split encoders and multi-agent DRL for joint optimization.	Cooperative multi-task processing; distributed environmental sensing.	Cooperation can be detrimental when tasks are unrelated; significant complexity in managing agent interactions and resources.
Explainable Systems	Transmits structured, human-interpretable SI (such as scene graphs, triplets), reducing data volume.	Moderate to High, due to extra processing for feature disentanglement, triplet extraction, or scene graph generation.	Tasks requiring trust, debugging, and verification, such as text QA or image retrieval.	Overcoming the "black box" nature of neural networks; potential trade-off between interpretability and raw task accuracy.
Generative/Diffusion Decoding	Generation-based decoding: the receiver synthesizes a perceptually convincing, task-appropriate output from a compact transmitted semantic description, reaching extreme compression (channel-bandwidth ratios down to 10^{-4}–10^{-3} for video).	High at the receiver, due to the iterative sampling process of diffusion models, which conflicts with low-latency requirements.	Image and video reconstruction/delivery judged by perceptual quality (Level B) or downstream task success (Level C).	Controllability/hallucination risk (generated output may not faithfully reflect the transmitted description); largely absent information-theoretic guarantees for generative decoders.
Semantic-aware Systems	Optimizes resource use by scoring semantic features by task importance and prioritizing the highest-ranked ones for transmission.	Moderate to High, due to the requirement for computing feature importance and solving complex resource allocation optimization problems.	Augmented Reality (AR); systems with user-defined quality ranges; tasks with subjective information value.	Accurately defining and measuring "semantic importance" is difficult; solving the resulting non-convex optimization problems is challenging.
RIS-assisted Systems	Improves the reliability of semantic transmission by reconfiguring the physical channel environment based on both channel conditions and the semantic importance of task-specific data streams.	Moderate, requiring joint optimization of the transmitter, receiver, and RIS, often via deep learning or RL, with additional complexity for task-aware phase optimization.	Enhances other ToSC applications, especially in hostile channel environments with low SNR or blockages; particularly beneficial for tasks with varying semantic importance across data streams.	Performance is fundamentally tied to the physical channel; requires complex, real-time control of RIS elements; additional complexity in jointly optimizing for channel conditions and task-specific semantic priorities.
ToSC in Edge Computing	Reduces wireless data volume and latency by extracting and transmitting SI at the network edge instead of raw data.	High, as it creates a joint communication-computation problem, adding processing load at both the edge device and the server.	Low-latency applications at the network edge; multimodal tasks using the edge as an intelligent semantic relay.	Requires careful optimization of the trade-off between communication and computation; introduces new resource management challenges.
Security & Privacy	Counters AI-native threats across the learning pipeline: feature obfuscation/disentanglement against model inversion, poisoning/backdoor detection and certified defenses, and adversarial training for robustness.	High, relying on adversarial training, feature disentanglement, certified smoothing, and dynamic encryption schemes.	Any task with sensitive data or safety-critical decisions vulnerable to model inversion, data poisoning, backdoor, adversarial-example, or large-model (e.g., fine-tuning-persistent backdoor) attacks.	Persistent trade-off between task utility/compute cost and protection; verifiable, certified guarantees remain rare; large-model and prompt-injection threats are largely unexamined.

References

[1] A. A. Alemi, I. Fischer, J. V. Dillon, K. Murphy, Deep variational information bottleneck, arXiv preprint arXiv:1612.00410 (2016).

[2] J. Shao, Y. Mao, J. Zhang, Learning task-oriented communication for edge inference: An information bottleneck approach, IEEE J. Sel. Areas Commun. 40 (1) (2021) 197–211. doi:10.1109/jsac.2021.3126087.

[3] E. Beck, C. Bockelmann, A. Dekorsy, Semantic information recovery in wireless networks, Sensors 23 (14) (2023) 6347. doi:10.3390/s23146347.

[4] J. Kang, H. Du, Z. Li, Z. Xiong, S. Ma, D. Niyato, Y. Li, Personalized saliency in task-oriented semantic communications: Image transmission and performance analysis, IEEE J. Sel. Areas Commun. 41 (1) (2022) 186–201. doi:10.1109/jsac.2022.3221990.

[5] S. Liu, N. Li, Y. Deng, T. Q. Quek, Goal-oriented semantic communication for wireless visual question answering, IEEE J. Sel. Areas Commun. 43 (12) (2025) 4247–4261, arXiv:2411.02452. doi:10.1109/jsac.2025.3623184.

[6] Z. Wang, Y. Deng, A. H. Aghvami, Goal-oriented semantic communications for avatar-centric augmented reality, IEEE Trans. Commun. (2024). doi:10.1109/tcomm.2024.3420708.

[7] W. Zhang, H. Zhang, H. Ma, H. Shao, N. Wang, V. C. Leung, Predictive and adaptive deep coding for wireless image transmission in semantic communication, IEEE Trans. Wireless Commun. 22 (8) (2023) 5486–5501. doi:10.1109/twc.2023.3234408.

[8] H. Xie, Z. Qin, G. Y. Li, B.-H. Juang, Deep learning enabled semantic communication systems, IEEE Trans. Signal Process. 69 (2021) 2663–2675. doi:10.1109/tsp.2021.3071210.

[9] J. Park, S. W. Yoon, Transmit what you need: task-adaptive semantic communications for visual information, IEEE J. Sel. Areas Commun. 43 (12) (2025) 4182–4197, arXiv:2412.13646. doi:10.1109/jsac.2025.3623159.

[10] E. C. Strinati, P. Di Lorenzo, V. Sciancalepore, A. Aijaz, M. Kountouris, D. Gündüz, P. Popovski, M. Sana, P. A. Stavrou, B. Soret, et al., Goal-oriented and semantic communication in 6G AI-native networks: The 6G-GOALS approach, in: Proc. 2024 Joint Eur. Conf. Netw. Commun. & 6G Summit (EuCNC/6G Summit), IEEE, 2024, pp. 1–6. doi:10.1109/eucnc/6gsummit60053.2024.10597087.

[11] Q. Hu, G. Zhang, Z. Qin, Y. Cai, G. Yu, G. Y. Li, Robust semantic communications with masked VQ-VAE enabled codebook, IEEE Trans. Wireless Commun. 22 (12) (2023) 8707–8722. doi:10.1109/twc.2023.3265201.

[12] Y. Blau, T. Michaeli, Rethinking lossy compression: The rate-distortion-perception tradeoff, in: Proc. Int. Conf. Mach. Learn., PMLR, 2019, pp. 675–685.