Tutorial on Semantic Communication for the Internet of Things: Frameworks, Optimization, and Open Challenges

Samer Lahoud, *IEEE Senior Member*

Faculty of Computer Science, Dalhousie University, Canada

1 Abstract

Semantic communication (SemCom) represents a shift from traditional bit-level transmission to the exchange of task-relevant meaning, with strong potential efficiency gains for future wireless networks. This tutorial provides a structured introduction to SemCom, starting from centralized architectures and then moving toward distributed, resource-constrained wireless Internet of Things (IoT) deployments. It reviews core principles and models, with a particular focus on Deep Joint Source–Channel Coding (Deep JSCC), and discusses the metrics commonly used to evaluate semantic encoders and decoders. Beyond this general survey, the tutorial is organized around three technical themes that are key to making SemCom practical for IoT. First, it examines distributed training of semantic encoders and decoders using federated and split learning, highlighting how data locality, heterogeneity, and wireless communication constraints shape these designs. Second, it presents adaptive runtime mechanisms, including multi-configuration Deep JSCC and semantic-aware control, that allow SemCom systems to react to changing channel conditions and device capabilities without full retraining. Third, it introduces semantic-aware resource allocation, where radio resources are managed with semantic quality, latency, and energy in mind rather than bit throughput alone. Attendees will gain a consolidated view of existing SemCom approaches, an understanding of how distributed learning and adaptation extend them to IoT scenarios, and a clear picture of open research directions.

2 Presenter

Dr. Samer Lahoud is an Associate Professor and University Research Chair in the Faculty of Computer Science at Dalhousie University, Canada. He is an IEEE Senior Member and currently serves as Head of the Systems Research Cluster at the Faculty of Computer Science. He holds a Ph.D. in computer science and networks from IMT Atlantique in Rennes and a Habilitation from the University of Paris-Saclay in France. Before joining Dalhousie, he was an Associate Professor at the University of Rennes and a researcher at IRISA, and he previously worked as a research engineer at Nokia Bell Labs Europe.

His research focuses on next-generation wireless networks, with emphasis on low-power wide-area networks, semantic communication, and AI-native wireless IoT systems. He develops analytical models, optimization methods, and data-driven algorithms for radio resource management, distributed learning, and performance evaluation.

3 Learning Outcomes

After this half-day tutorial, participants will be able to:

- Explain the core concept of semantic communication (SemCom) and clearly distinguish it from bit-centric communication in terms of objectives, interfaces, and evaluation.
- Recognize IoT scenarios where SemCom provides practical value, by mapping application goals (task success, latency, energy, reliability) to sensing modalities and traffic patterns.
- Choose suitable metrics and evaluation workflows for SemCom-over-wireless experiments, including task-oriented success measures and robustness under channel and device heterogeneity.

- Compare the main system design options for learning and deployment in IoT, including centralized training versus distributed learning (federated and split learning), and reason about trade-offs in overhead, privacy, and device constraints.
- Outline how semantics can be integrated into wireless resource management, by formulating semantic-aware objectives and constraints for scheduling and link adaptation, and identifying key open problems for research and standardization.

4 Description

The tutorial is organized in four parts that move from centralized SemCom models to distributed, resource-constrained IoT deployments. Each part combines a compact survey of existing work with design insights on distributed training, adaptive runtime operation, and semantic-aware resource allocation.

4.1 Semantic communication foundations and Deep JSCC for IoT

This part introduces SemCom from a system perspective, starting from centralized architectures and then motivating IoT deployments.

- From classical source–channel separation to semantic and task-oriented communication, and when semantic approaches provide advantages over bit-level transmission.
- Conceptual definition of SemCom and distinction between reconstruction-oriented Deep Joint Source–Channel Coding (Deep JSCC) and task-oriented SemCom.
- Survey and classification of SemCom approaches by objective (distortion, task accuracy, hybrid) and modality (images, text, speech, sensor data).
- Deep JSCC architectures: encoder–channel–decoder autoencoders, typical loss functions, and illustrative examples for image and sensor data.
- Assumptions of centralized SemCom models (global datasets, powerful hardware, idealized channels) and why IoT, with its task orientation, sporadic traffic, and tight constraints on bandwidth, latency, and energy, requires new frameworks.

4.2 Training and operating semantic encoders in IoT: federated, split, and adaptive models

This part focuses on distributed training and adaptive runtime operation of SemCom models in heterogeneous IoT environments.

- IoT constraints for learning and inference: data locality, statistical and systems heterogeneity, intermittent connectivity, and strict energy budgets.
- Federated learning for SemCom: central ideas, heterogeneity-aware clustering and client selection, and trade-offs between semantic fidelity, communication overhead, and fairness across devices.
- Split learning and split inference: partitioning semantic encoders between devices and edge servers, static versus adaptive cut selection, and impact on latency, energy consumption, and privacy.
- Adaptive runtime mechanisms: multi-configuration or slimmable Deep JSCC models, SNR-conditioned encoders, and semantic-aware control that selects configurations based on channel state and device capabilities without full retraining.

4.3 Semantic-aware resource allocation for SemCom IoT

This part addresses radio resource management when semantic quality, latency, and energy are primary objectives.

- Limitations of bit-centric schedulers and link adaptation when applied directly to SemCom-enabled IoT systems.

- Definition of semantic utility that combines semantic fidelity, latency, and energy consumption, and its role in decision making.
- Design patterns for semantic-aware scheduling: prioritizing high-value semantic updates under congestion, mapping feature importance to time–frequency resources, and coordinating multiple semantic flows.
- High-level use of optimization and reinforcement learning for semantic-aware radio resource management, and relation to existing work on learning-based scheduling and control in wireless networks.

4.4 Open challenges and research directions in SemCom for IoT

The final part synthesizes the tutorial and highlights open problems that structure future research.

- Reliability and guarantees for semantic channels, including semantic outage definitions, performance bounds, and interaction with classical quality of service metrics.
- Multi-modal SemCom, privacy and security issues, and robustness to adversarial behaviour and poisoned or unreliable devices.
- Fairness and inclusion in distributed semantic training and scheduling across heterogeneous IoT nodes, including bias in data and resource allocation.
- Experimental challenges and standardization aspects on the path from laboratory prototypes to AI-native wireless IoT deployments.

5 Potential Audience

This tutorial is intended for researchers, graduate students, and professionals working in wireless communications, networking, and data-driven or AI-based systems who wish to understand how semantic communication can be brought from centralized prototypes to practical Internet of Things deployments. It will be particularly relevant for participants involved in the design of future 5G/6G and IoT platforms, edge intelligence, and learning-enabled communication protocols, who are interested in how semantic encoders, distributed training, and semantic-aware resource allocation fit into existing wireless architectures. A general familiarity with modern wireless networks and machine learning is helpful, but the material is structured to remain accessible to a broad audience and will emphasize concepts, design patterns, and system-level insights rather than detailed mathematical derivations.

6 Related Events

This tutorial was previously delivered in September 2025 as an online tutorial hosted by the IEEE Canadian Atlantic Section (CAS), under the title *Semantic Communications for IoT*. The session attracted approximately 30 attendees, with a mixed audience of graduate students, academic researchers, and industry participants. It was delivered in a lecture-style format with dedicated time for questions and discussion.

In addition, Dr. Lahoud has delivered several tutorials at international conferences and professional forums, including:

- Lahoud, S.; El Helou, M. (2019). Tutorial on Low Power Wide Area Networks for the Internet of Things: Framework, Optimization, and Challenges of LoRaWAN and NB-IoT. 22nd International Symposium on Wireless Personal Multimedia Communications (WPMC 2019), Lisbon, Portugal.
- Lahoud, S.; El Helou, M. (2019). Tutorial on Cellular Internet of Things: Framework, Optimization, and Challenges of NB-IoT. IEEE 5G World Forum, Dresden, Germany.
- Lahoud, S.; El Helou, M. (2018). Tutorial on Low Power Wide Area Networks for the Internet of Things: Framework, Performance Evaluation, and Challenges of LoRaWAN and NB-IoT. 25th International Conference on Telecommunications (ICT 2018), Saint-Malo, France.
- Lahoud, S. (2018). Tutorial on Low Power Wide Area Networks for the Internet of Things: Framework, Optimization, and Challenges. International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS 2018), Bordeaux, France.

7 Novelty

Semantic communication has rapidly emerged as a key research direction in the context of AI-native 6G networks. Recent work on Deep Joint Source–Channel Coding and task-oriented communication has demonstrated that learning-based transceivers can significantly improve robustness and spectral efficiency when the objective is task performance rather than bit-perfect reconstruction. As a result, SemCom now appears in 6G roadmaps and white papers as a candidate technology for future wireless systems, especially in scenarios involving rich sensing data and intelligent edge processing.

At the same time, most existing SemCom prototypes are developed in centralized, resource-rich settings: models are trained on global datasets, executed on powerful hardware, and evaluated under idealized channel assumptions. This stands in sharp contrast with wireless Internet of Things deployments, where data is distributed across many devices, energy and computation budgets are limited, connectivity is intermittent, and radio resources must serve a very large number of nodes. Moving SemCom from centralized laboratories to distributed, resource-constrained IoT networks therefore raises fundamental challenges in model training, runtime adaptation, and resource allocation.

This tutorial focuses on these challenges and is organized around the following guiding questions:

- Which architectural and training choices enable SemCom under wireless IoT constraints such as limited energy, bandwidth, and intermittent connectivity?
- How can federated and split learning be adapted to train semantic encoders and decoders over wireless networks while respecting data locality and heterogeneity?
- How can SemCom systems adapt at runtime to varying channel conditions, device capabilities, and traffic patterns using multi-configuration models and semantic-aware control?
- How should radio resources be allocated when semantic quality, latency, and energy consumption are primary objectives, rather than bit throughput alone?

By explicitly tracing the path from centralized SemCom models to distributed IoT deployments, the tutorial aims to provide attendees with a coherent view of the field, a set of reusable building blocks, and a structured perspective on the most pressing open research directions.

8 Additional Infromation

Attendees will receive access (via a public repository) to the tutorial slide deck with diagrams, key formulas, literature references, and links to public code repositories, along with a concise concept sheet summarizing core SemCom notions and building blocks, and pseudo-code workflows for Deep JSCC training, a federated loop under wireless constraints, and runtime configuration selection. Lightweight hands-on material will also be provided, including compact Python/PyTorch notebooks that implement a simple Deep JSCC pipeline over an AWGN channel and a minimal federated training loop, together with illustrative plots comparing reconstruction quality and task performance against classical schemes across SNR regimes and brief instructions for running and extending these notebooks.