

当生成式AI遇上语义通信：LLM如何让6G网络从“信息恢复”迈向“信息再生”

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/42743.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

当生成式AI遇上语义通信：LLM如何让6G网络从“信息恢复”迈向“信息再生”。 香港中文大学（深圳）崔曙光教授团队联合鹏城实验室、香港大学等研究人员在中国工程院院刊Engineering上发表题为Generative Semantic Communication: Architectures, Technologies, and Applications（生成式语义通信：架构、技术与应用）的综述性研究文章。该研究系统梳理了变分自编码器、生成对抗网络、扩散模型三类生成式AI模型驱动的语义通信系统，并创新性提出基于大语言模型（LLM）的原生生成式语义通信完整架构，为6G无线网络中多模态轻量化语义传输提供了系统性理论基础与技术实现路径。

传统比特通信范式存在哪些固有局限？

5G及前代无线网络以比特传输指标为核心设计思路，依托香农通信理论完成原始数据的完整收发。然而，海量物联网、无人机、车载终端产生的图像、语音、视频等多模态数据，使得完整的比特传输产生巨大带宽开销，难以满足未来全息通信、自动驾驶等低时延大流量场景的需求。

语义通信以传输语义信息替代原始比特，已成为6G关键技术方向。然而，现有的基于判别式神经网络的语义通信系统面临三重瓶颈：跨场景泛化能力不足、复杂噪声信道下重建质量大幅下降、缺乏高层语义推理能力。生成式人工智能（GAI）能够学习复杂数据分布，实现高低维数据转换，有效弥补了上述短板，为语义通信开辟了全新路径。

三类基础生成模型如何支撑语义通信？

研究团队依次介绍了变分自编码器（VAE）、生成对抗网络（GAN）以及扩散模型（DM）三类经典生成模型驱动的语义通信系统，分别阐述了各模型基础原理、收发端编解码部署形式与现有研究成果。

1、VAE驱动语义通信

变分自编码器属于概率生成模型，其中概率编码器将输入数据压缩至低维隐空间，概率解码器基于含噪隐向量重建原始数据。分层VAE、条件VAE、向量量化VAE等衍生模型已被广泛应用于远程控制、文本语音、短包图像传输等场景。基于VAE的语义通信系统架构如图1所示，其中输入数据经过概率编码器生成低维隐向量，通过无线信道传输后，由概率解码器完成数据重建，从而实现先压缩后重建的高效语义传输。

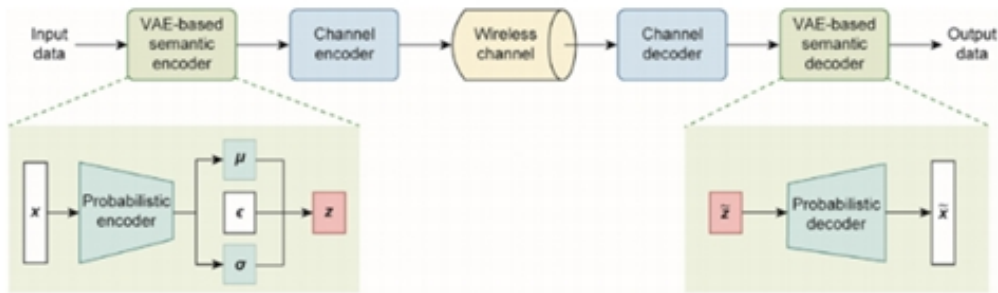


Fig. 1. VAE-enabled SemCom architecture. x : input data; μ : mean; σ : standard deviation; ϵ : random vector; z : latent vector; $\tilde{z}$: noisy latent vector; $\tilde{x}$: reconstructed data.

47

图1 VAE驱动的语义通信系统架构

2、GAN驱动语义通信

生成对抗网络由生成器、判别器构成对抗训练架构。发送端通过反演机制提取数据语义特征，接收端生成器基于含噪语义特征完成数据重建。现有研究借助条件GAN、CycleGAN等变体优化隐私保护、多模态适配等能力。基于GAN的语义通信系统架构如图2所示，其中生成器与判别器的对抗博弈使接收端能够生成高度逼真的数据内容。

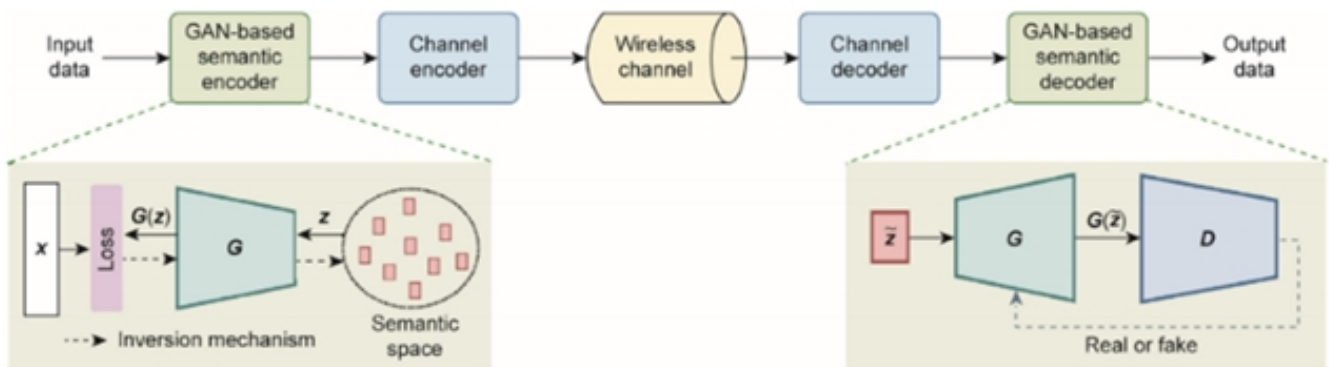


Fig. 2. GAN-enabled SemCom architecture. G : generator; D : discriminator; $G(z)$: synthetic data; $G(\tilde{z})$: synthetic data based on $\tilde{z}$.

图2.GAN驱动的语义通信系统架构

3、扩散模型驱动语义通信

扩散模型包含前向加噪、反向去噪两大过程，可同时完成语义编码、信道建模均衡、语义解码三项功能，天然适配图像、语音、三维元宇宙内容传输场景。基于扩散模型的语义通信系统架构如图3所示，其逐步去噪过程对信道噪声具有极强鲁棒性，能够有效抑制无线传输中的干扰。

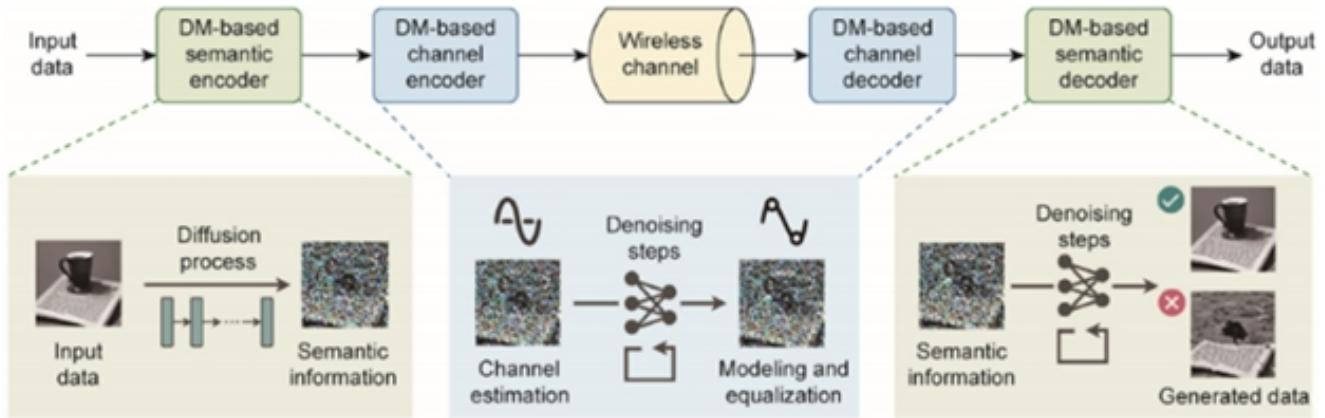


Fig. 3. DM-enabled SemCom architecture.

图3.扩散模型驱动语义通信系统架构

上述三类生成模型虽各具优势，但均仅适配单一模态与单一任务场景，无法支撑多模态、多任务一体化语义传输。为此，研究团队提出了原生大模型的生成式语义通信架构，实现一种架构支撑所有场景的通用设计。

原生LLM生成式语义通信系统如何实现信息再生？

研究团队提出了一个基于LLM的原生生成式语义通信架构，整套系统由发送端LLM智能体、接收端LLM智能体以及双向信道适配模块构成完整传输链路，总体架构如图4所示。该架构采用全新传输逻辑：发送端提取轻量化理解嵌入完成传输，接收端无需还原原始输入数据，而是依托理解嵌入直接生成目标业务内容——通信范式由此从传统的信息恢复跃迁为信息再生。

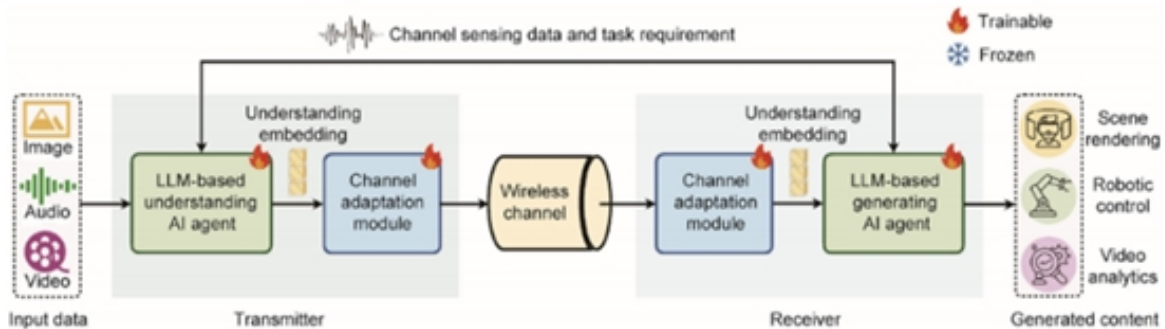


图4.原生LLM生成式语义通信整体架构

整套系统包含四个核心模块：

双端专用大模型智能体：收发两端分别部署理解智能体与生成智能体，发送端完成多模态数据语义提取并生成轻量化理解嵌入，接收端依托理解嵌入直接生成目标业务内容，摒弃传统比特重建思路；两类智能体依靠专属角色提示词区分功能，实现端到端协同。

多模态感知编码器：将图像、音频、视频等多模态输入统一映射至大模型兼容的token特征空间，消除模态鸿沟，真正实现任意模态输入-统一语义传输-任意任务输出的目标。

轻量化微调模块：冻结预训练大模型主干参数，仅微调外层适配模块，大幅降低边缘终端部署的算力开销，使大模型在资源受限设备上依然可落地。

记忆与工具集成单元：短期记忆维持通信上下文连贯，长期记忆沉淀通用语义知识；配套检索、检测工具拓展复杂任务处理能力，使智能体具备持续进化能力。

整套系统采用语义嵌入传输+本地内容再生的设计思路，大幅降低传输比特量，同时自适应无线信道时变特性，为6G多场景语义通信提供了统一解决方案。

视频检索场景测试：精度提升53%，开销骤降99.98%

研究团队搭建了点对点视频检索仿真环境，以视频目标检索作为测试任务，选取YOLOv8-DeepSORT、InternVL-1.5、Qwen等模型完成系统配置，将所提的生成式语义通信方案与传统MPEG+LDPC传输方案、基于VAE/GAN/DM的三类生成式基线方案进行对比，实验观测到下面三组结果。其中，检索精度对比见图5，各类方案通信开销见表1。

检索精度：原生LLM生成式方案平均检索精度达93.03%，相比传统MPEG+LDPC传输方案提升53%，大模型强大的语义理解与生成能力功不可没；

信道鲁棒性：原生LLM生成式方案精度标准差更小，在不同信噪比下性能波动更低，抗信道噪声能力显著优于其他方案，这得益于大模型先验知识可有效补偿噪声影响；

通信开销：原生LLM生成式方案仅需36kb传输量，传统方案通信开销为219Mb，传输数据量下降99.98%，真正实现极致压缩、精准生成。

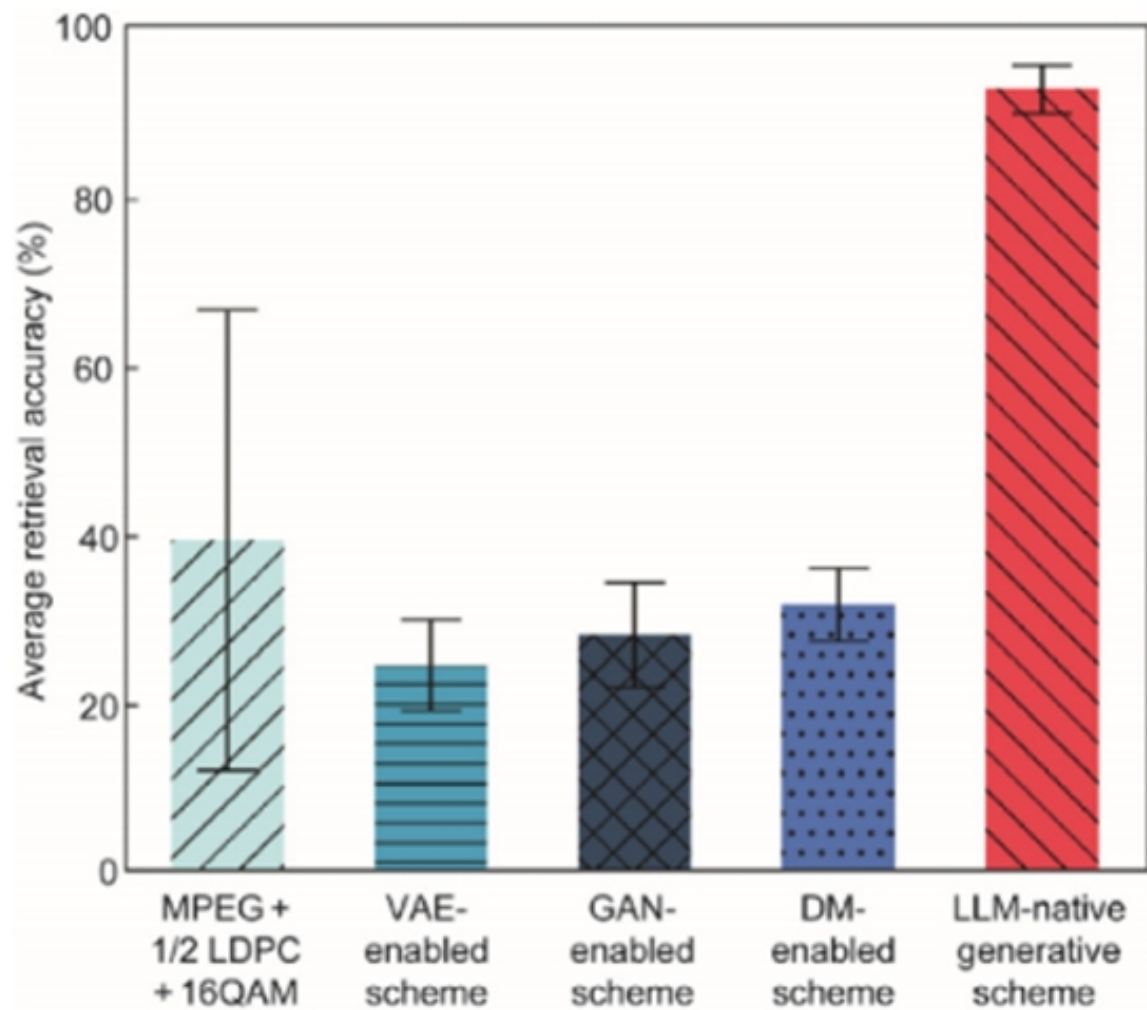


图5.不同方案检索精度对比

表1各类方案通信开销对比

Communication overhead of the five schemes.

Scheme	Communication overhead
MPEG + 1/2 LDPC + 16QAM	219 Mb
VAE-enabled scheme	13 Mb
GAN-enabled scheme	13 Mb
DM-enabled scheme	13 Mb
LLM-native generative scheme	36 kb

从工业物联到低空经济，生成式语义通信能落地何处？

研究团队发现生成式语义通信在四类场景中展现出跨场景适配潜力。在工业物联网中，设备发送语义指令替代原始数据，接收端生成操作决策，降低信令开销、提升协同效率；在车联网V2X中，车辆交换语义路况信息并实时生成驾驶指令，兼顾低时延与隐私安全；在元宇宙中，用户动作与场景描述实现轻量化语义传输，接收端实时生成虚拟内容，兼顾沉浸体验与数据隐私；在低空经济中，无人机群将感知数据浓缩为语义摘要，赋能地面平台生成调度指令，缓解空域拥塞。各场景均可提取专属语义信息实现轻量化传输，兼顾通信效率与隐私安全，展现出从工业制造到智慧交通、从虚拟交互到低空飞行的泛在服务能力。

本文提出了全新的原生LLM生成式语义通信系统。依托收发端双大模型智能体架构，推动通信范式由信息恢复向信息再生转变，并通过视频检索仿真实验，从检索精度、信道鲁棒性、通信开销三个维度验证了方案的显著性能优势。

然而，当前技术落地仍面临三项关键挑战：一是大模型参数量巨大，边缘终端算力与存储不足，轻量化部署难度较高；二是无线信道动态变化、业务类型复杂，收发两端智能体易产生知识偏差，双向协同机制有待完善；三是大模型存在训练数据记忆特性，语义传输过程易引发数据推理、投毒攻击等安全隐患，完整安全传输体系仍有缺失。

总体而言，生成式人工智能与语义通信的深度融合，是6G无线网络的重要发展方向。未来研究可从模型轻量化（量化、剪枝、蒸馏）、智能体持续协同学习、语义传输安全机制三条路径推进，为下一代语义通信系统设计与工程落地提供系统性理论基础。当大模型成为收发端的大脑，通信将不再是比特的搬运，而是语义的共鸣——一个极简传输、极致智能的6G新纪元，正悄然走来。（来源：EngineeringJournals微信公众号）

相关论文信息：<https://www.engineering.org.cn/engi/EN/10.1016/j.eng.2025.07.022>

作者：崔曙光等 来源：《工程》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发