东吴证券 SOOCHOW SECURITIES

汽车行业深度报告

AI 系列深度 04 期：什么是表征？

2026 年 07 月 29 日

增持（维持）

证券分析师 黄细里
执业证书：S0600520010001
021-60199793
huangxl@dwzq.com.cn

投资要点

- **表征是模型对现实世界信息形成的内部抽象表示，是人工智能进行理解、推理与预测的基础。**表征学习质量直接影响下游任务上限，Bengio、Courville、Vincent 等指出，机器学习效果很大程度取决于数据表示方式。我们将表征拆为三层：数据层的向量编码、模型层的中间特征、哲学层的意义承载问题。
- **表征技术演进可概括为四个阶段：**符号 AI 依赖人工设计的离散符号和逻辑规则；特征工程依赖 SIFT、HOG、TF-IDF 等专家特征与浅层模型；深度学习通过神经网络端到端学习连续表征，AlexNet 与 Transformer 是关键节点；基础模型进一步走向跨任务、跨模态统一表征空间，GPT-3、CLIP 等是代表。AI 范式迁移的底层线索，是机器记录与组织世界方式的变化。
- **围绕表征如何获得和承载，当前存在四组路线分歧：**表征中心与生成中心，分别强调先学好表示或通过生成任务形成能力；符号表征与分布式表征，对应符号主义、连接主义及神经符号融合；离散表征与连续表征，对应 token 化和潜空间路线；显式世界模型与隐式潜空间，尤其在具身智能中体现为是否需要生成可见未来状态。
- **表征是否等于理解仍是开放问题。**中文房间、符号接地、随机鹦鹉等观点质疑模型是否真正理解，柏拉图表征假说等研究则提示不同模型表征可能趋同。此外，表征坍缩、可辨识性、好表征判据、多模态统一表征能否成立，均是当前技术和哲学交叉的重要议题。
- **落到物理 AI，表征视角可解释自动驾驶与机器人路线分歧。**自动驾驶围绕一段式与两段式、端到端与 VLA、世界模型云端造数据与车端决策、显式生成与隐式潜空间展开；机器人则进一步面对动作离散 token 还是连续轨迹、数据来自人类视频、真机还是仿真、单体 VLA 还是快慢双系统等问题。两者共性在于都要编码物理世界，差异在数据来源、实时性与安全闭环。
- **我们认为,数字 AI 底座模型发展路径已逐渐清晰、价值在于确定性，但其难以实现物理世界的建模闭环；我们认为物理 AI 将成为物理世界的 AI 解决方案，从而在 AI 的当前发展阶段成为增量更大的方向，智能驾驶作为最先跑通闭环的代表场景，我们认为应重点关注。**
- **风险提示：**技术迭代不及预期的风险，大模型厂商 ARR 增速放缓的风险，云服务商资本开支不及预期的风险，智能驾驶及机器人商业化落地不及预期的风险。

行业走势

相关研究

《AI 系列深度 03 期：数字 AI 与物理 AI 的边界在哪里？》

2026-07-28

《AI 系列深度 02 期：AGI 是什么？》

2026-07-28

内容目录

图表目录

1. 表征：人工智能理解世界的底层语言

表征是理解 AI 技术演进的底层变量。人工智能要完成识别、生成、推理或控制，首先需要把现实世界编码为机器可计算的内部表征。后续关于嵌入与自监督、CNN/ViT、RNN/Transformer、生成模型的讨论，本质上都围绕“如何获得更有效的表征”展开。

表征可以理解为机器对世界的内部记法，指 AI 系统将图像、文本、声音、动作等外部信息转化为可计算编码的过程。文本 token 与图像 patch/token 属于数字表征，不能直接等同于自动驾驶中的三维物理世界表征；但二者共同指向同一问题，即外部信息如何进入模型并被用于后续计算。在机器学习中，这类编码通常以向量形式存在。

分布式表征强调由多个单元共同表示一个对象，单元与对象之间形成多对多关系。Hinton、McClelland、Rumelhart 在 1986 年提出的分布式表征思想，提供了连接主义学习的基础：新对象即使未在训练集中逐一出现，只要与已见对象在特征空间中接近，也可以获得有意义的表示。AI 表征可以视为神经活动模式在数学空间中的对应形式。

表征的重要性在于，它直接决定模型能力边界。Bengio、Courville、Vincent 在 2013 年表征学习综述中指出，机器学习算法能否成功，很大程度取决于数据以何种方式被表示。识别、生成、推理、决策等任务的上限，均受到表征质量约束；因此，表征是理解 AI 技术演进与路线分歧的主线。

2. 表征三层含义：数据编码、模型特征与意义问题

“表征”在不同语境下对应不同层次，可将其拆分为三层：（1）数据层回答原始输入如何转化为数字编码；（2）模型层回答神经网络内部形成了哪些中间特征；（3）哲学层回答这些向量是否能够承载真实含义。三层分别对应输入编码、内部特征与意义问题。

表1：“表征”一词的三层含义

层面	回答的问题	典型对象	代表性议题
数据层面	怎样把原始输入编码成机器可算的形式	像素矩阵、token、embedding	输入表征的选择如何影响下游性能
模型层面	神经网络内部学到了什么	隐藏层激活、潜空间	表征的可解释性与可控性
哲学层面	这些向量是否“代表”了真实含义	符号与意义的关系	统计相关性是否等于真正理解

数据来源：《Representation Learning: A Review and New Perspectives》等其他文献，东吴证券研究所

2.1. 数据层面：把原始输入编码成向量

从原始信号到向量，是 AI 系统完成计算的起点。不同输入对应不同编码方式：图像通常先形成像素矩阵并提取边缘、纹理等特征；文本先切分为 token，再映射为 embedding；药物分子可由 SMILES 字符串或图结构表达。对物理 AI 而言，表征还需要覆盖三维空间、动态交互、动作后果与可行动状态。表征方式决定信息保留与损失，进而影响下游任务上限。

词向量是数据表征的经典案例，词被映射为向量后，词义关系可在向量空间中体现为方向关系，典型例子是“国王-男人+女人≈王后”，即从 “国王” 中减去 “男性” 这个属性，得到 “王室 / 统治者”的抽象概念向量，再加上“女性”属性，就得到“王后”。好的表征能够将语义相似性转化为空间邻近性，使含义接近的对象在数字空间中保持接近。

输入表示方式直接影响模型效果。同一对象采用不同编码，模型可达到的性能上限也会变化。Bengio 等指出，表征学习的价值在于尽量拆解数据背后相互纠缠的因素，即解缠（disentangle）使后续任务更易学习。落到工程侧，分词粒度、图像切块方式、分子描述形式等选择，往往与模型规模共同决定最终表现。

2.2. 模型层面：神经网络中间层学到了什么

隐藏层激活模式构成模型学习到的内部表征。深度网络将原始输入逐层加工为更抽象的特征；Hinton 与 Salakhutdinov 在 2006 年证明，自编码器可以将高维数据压缩为低维 codes，效果优于传统主成分分析。这组压缩编码所在空间即潜空间是模型组织信息与生成内容的重要载体。

潜空间承担压缩与编辑两类功能。在潜空间中，语义相近的对象距离更近，关系也可表现为几何结构。扩散模型等生成模型通常先将图像压缩到潜空间，在低维空间完成去噪、修改或生成，再还原为图像，从而降低计算成本。Rombach 等提出的隐扩散模型是该路径的代表。

可解释性问题的核心在于，模型内部表征能否被识别和调控。由于表征由数据驱动形成，研究者通常难以直接读取其含义，因此发展出探针等工具，用于检验特定网络层编码了哪些信息、能否被人为干预。这一问题与后文关于好表征、可控表征和产业安全性的讨论直接相关。

2.3. 哲学层面：表征能否承载真实含义

“理解”问题构成表征研究的哲学分歧。一类观点认为，智能行为需要内部表征结构支撑，理解概念意味着系统内部形成对应表达；另一类观点更强调外部行为效果，对内部是否存在真实意义保持审慎。大模型在多任务上的表现提升，使“模型内部向量是否真正承载含义”成为重新被讨论的核心问题。

统计相关并不必然等于真正理解。Searle 的中文房间思想实验指出，系统即便能够

按照规则处理符号，也不代表其理解了符号含义。Bender 等提出“随机鹦鹉”质疑，认为语言模型可能只是复现高频共现关系，而非与真实世界建立稳定指向。这些问题关系到模型能力边界、可解释性与安全性的判断。

3. 表征演进：从人工符号到统一基础模型

表征演进的主线，是从人工手写走向机器自学，从孤立符号走向连续向量，并进一步走向跨模态统一空间。我们将其概括为四个阶段。每一次 AI 范式迁移，底层都伴随机器记录、压缩和组织世界方式的变化。

表2：AI 表征演进的四个时代

时代	代表方法	表征特点	局限性
符号 AI 时代，1956—1990s，	知识图谱、逻辑规则、专家系统	人工设计、离散、可解释	无法处理模糊性与不确定性
特征工程时代，1990s—2012，	SIFT、HOG、TF-IDF、词袋	人工设计特征 + 浅层学习	依赖专家经验，迁移性差
深度学习时代，2012—2020，	CNN、RNN、Transformer	端到端自动学习连续表征	需大量标注数据，领域专用
基础模型时代，2020—今，	GPT、CLIP、多模态大模型	统一表征空间、跨模态对齐	可解释性、幻觉与对齐问题

数据来源：《ImageNet Classification with Deep Convolutional Neural Networks》，《Attention Is All You Need》等其他文献，东吴证券研究所

3.1. 符号 AI 时代：人工规则构建可解释表征

早期 AI 将知识表征为符号和规则。Newell 与 Simon 提出物理符号系统假设，主张能够操作符号的系统具备产生智能的充分条件。该路线的优势在于过程透明、推理链条清晰、可解释性较强；局限在于知识需要人工编码，难以覆盖真实世界中的模糊性与例外情况。

符号 AI 的核心约束是接地困难和扩展成本较高。Harnad 提出符号接地问题，指出符号本身只是形式，系统需要将其与感知和行动经验建立联系，才能获得稳定含义。该问题推动后续研究转向从数据中自动学习表征，而不是完全依赖人工规则。

3.2. 特征工程时代：专家特征支撑浅层学习

特征工程阶段由专家决定模型应关注哪些信息。图像任务中，SIFT、HOG 等方法提取边角、轮廓、明暗变化等视觉线索；文本任务中，TF-IDF 通过词频和逆文档频率衡量词项区分度。人工特征再输入浅层分类器，形成“专家设计表征、模型完成学习”的典型范式。

该阶段的主要局限在于迁移性和扩展性不足。特征质量高度依赖专家经验，且针对A任务设计的特征迁移到B任务后往往效果下降。随着数据规模和算力提升，人工特征逐渐成为性能瓶颈，深度网络自动学习表征成为更具扩展性的路径。

3.3. 深度学习时代：神经网络自动学习表征

AlexNet验证了深度网络自动学习表征的可行性。2012年，Krizhevsky、Sutskever与Hinton提出的AlexNet显著降低ImageNet图像识别错误率，证明深度网络可以端到端学习层层递进的视觉特征，不再依赖专家手工设计中间特征（NIPS 2012）。此后，CNN主导视觉任务，RNN主导序列任务，表征开始大规模由数据驱动生成。

Transformer提供了更通用的表征骨架。Vaswani等在2017年提出基于注意力机制的Transformer架构，通过动态建模序列中不同位置之间的相关性，使模型能够跨长距离建立依赖关系。该机制随后从语言扩展到图像和多模态任务，成为统一表征的重要基础。

这一代表征虽然可以自主学习，但大多围绕单一任务训练，依赖大量人工标注，跨任务与跨模态迁移能力有限。更大规模的预训练与统一学习目标由此成为后续突破方向。

3.4. 基础模型时代：统一表征支撑多任务迁移

自监督与大规模预训练降低了人工标注依赖。BERT通过掩码语言建模进行预训练，再通过微调适配多类下游任务；GPT-3通过大规模自回归训练展示出少样本任务适配能力。自监督训练的核心，是利用数据自身构造学习信号，使模型在未显式人工标注的情况下形成可迁移表征。

跨模态统一表征推动图文进入同一语义空间。CLIP使用大规模图片和文本配对数据训练，使图像与文字能够在统一空间中对齐，并支持零样本分类。斯坦福随后以基础模型概括这类以大数据、大规模训练为基础、可适配多类下游任务的模型形态。

统一表征提升了通用性，也放大了可解释、幻觉与对齐问题。模型内部机制难以直接读取，生成内容可能偏离事实，价值偏好也需要进一步约束；这些问题与“随机鹦鹉”式理解争议相互交织，构成后续路线分歧和核心争议的现实背景。

4. 表征路线分歧：生成、符号、离散与世界模型

基础模型时代的路线分歧，主要围绕表征如何获得、由什么形式承载展开。我们归纳为四组问题：1）表征是专门学习，还是在生成任务中涌现；2）知识由符号规则承载，还是由分布式向量承载；3）表征应离散化为token，还是保持连续潜空间；4）物理决策是否需要显式生成未来画面。

4.1. 路线一：表征中心 vs 生成中心

表征中心路线主张先优化表征，再迁移到下游任务。对比学习通过拉近同一对象的

不同视图、拉远不同对象来学习稳定特征，CLIP、SimCLR、MoCo 属于该路径；掩码自监督则通过遮挡输入并要求模型还原内容来形成表征，MAE 是代表方法。该路线的共同点是显式优化“认识世界”的能力。

生成中心路线主张表征可以在生成任务中形成。GPT 通过预测下一个词学习语言分布，扩散模型通过逐步去噪学习图像或数据分布，Ho 等和 Rombach 等的工作是典型代表。该路线强调，模型在学习生成数据的过程中，也会内生形成可用于识别、推理和迁移的表征。

争论焦点在于，好的表征应当来自专门优化，还是来自生成过程中的涌现。学界常以生成式与对比式、判别式自监督进行对照。较为中性的理解是两者互补：对比式更强调判别稳定性，生成式更强调分布建模和补全能力，不同场景下最优路径并不相同。

表3：表征中心与生成中心路线对比

维度	表征中心	生成中心
核心目标	直接学习可迁移的表征	建模数据分布／生成数据
代表方法	CLIP、SimCLR、MoCo、MAE	GPT 自回归、DDPM、隐扩散
训练信号	对比损失、掩码重建	下一个词／逐步去噪
表征来源	显式优化表征目标	生成过程中涌现
机制特征	先优化可迁移特征	在生成分布中形成表征

数据来源：《Learning Transferable Visual Models From Natural Language Supervision》，《High-Resolution Image Synthesis with Latent Diffusion Models》等其他文献，东吴证券研究所

4.2. 路线二：符号表征 vs 分布式表征

符号路线强调知识以清晰的符号和逻辑规则表达，从而支持组合推理和可解释输出。Fodor 与 Pylyshyn 提出系统性质疑，认为人的思维具有组合规律，若神经网络不能解释这种组合性，则其内部可能仍需要某种符号结构支撑。

连接主义路线强调知识分布在网络参数和激活模式中。Hinton 等提出的分布式表征为该路径提供基础，深度学习和大模型进一步证明，在无显式规则的情况下，模型也可以通过数据和参数更新形成复杂能力。其优势在于可扩展，约束在于解释性较弱。

神经符号路线尝试结合两类机制：感知与表征由神经网络承担，规则推理和证明由符号系统完成。Marcus 长期呼吁混合路线；DeepMind 的 AlphaGeometry 用神经网络提出辅助线，再由符号引擎完成严格证明，在奥数几何题上接近金牌选手水平。

4.3. 路线三：离散表征 vs 连续表征

离散路线将连续信号映射为有限集合中的 token。VQ-VAE 先构建码本再将输入匹

配为最接近的码本编号；后续 VQ-VAE-2、VQGAN 将图像压缩为离散视觉 token，并结合 Transformer 建模。该路线的优势在于可复用语言模型处理 token 序列的成熟框架。

连续路线保留输入在连续数字空间中的表示。标准 embedding 和连续潜空间不将信号切分为离散编号，而是在连续空间中完成表示、编辑和生成；隐扩散模型即在连续潜空间中完成图像生成。该路线通常具备信息损失较低、轨迹更平滑等优势。

争论焦点在于，潜空间应离散化为 token，还是保持连续表示。LeCun 批评逐 token 预测容易产生误差累积，且现实世界信号并不天然以离散块存在，因此主张 JEPA 在压缩后的连续表征空间中预测，而非逐块生成。离散 token 路线更贴合现有大模型工程体系，连续潜空间路线则强调物理世界建模和高维连续信号处理效率。

4.4. 路线四：显式表征 vs 隐式表征

显式路线主张先生成可观察的未来状态，再用于决策。在机器人和自动驾驶中，世界模型可以生成未来画面或视频，辅助模型评估行动后果。Wayve GAIA-1、DeepMind Genie 和 OpenAI Sora 等工作，均体现了通过视频或交互世界建模来学习环境动态的方向。该路线的核心假设是，可见未来的生成有助于提升决策稳健性。

隐式路线主张不必生成像素级画面，而是在潜空间预测未来关键状态。LeCun 的 JEPA 和 Meta V-JEPA 系列聚焦语义特征预测；Dreamer 系列也在潜空间中进行规划和动作学习。该路线强调抽象状态预测更节省算力，也更贴近决策所需信息。

争论焦点在于，物理 AI 决策是否需要生成可见未来画面，还是只需要预测潜空间中的关键状态。

表4：四组表征路线之争一览

派系	显式路线	隐式路线	争论焦点
学习目标之争	表征中心，CLIP、MAE，	生成中心，GPT、扩散，	表征是专门学出来的还是生成中涌现的
范式之争	符号表征	分布式表征，连接主义，	大模型是否涌现真正的符号推理
表征形式之争	离散表征 VQ-VAE	连续表征，连续潜空间、JEPA，	潜空间采用离散 token 还是连续表示
世界模型之争	显式世界模型	隐式潜空间世界模型	决策是否需要先生成看得见的未来

数据来源：《A Path Towards Autonomous Machine Intelligence》，《Neural Discrete Representation Learning》等其他文献，东吴证券研究所

5. 核心争议：理解、坍缩与统一表征仍待验证

表征研究仍面临五类基础问题：理解边界、坍缩机制、可辨识性、好表征标准与多模态统一表征。

5.1. 表征是否等于理解

大模型学到的表征是否等同于理解，是最根本的争议之一。除中文房间和符号接地问题外，Bender 与 Koller 提出章鱼实验，强调模型即使能够通过文本规律完成对话，也可能没有将语言与真实世界建立联系。因此，仅基于文本共现形成的表征，不必然等同于面向真实世界的理解。

5.2. 表征坍缩问题

表征坍缩指模型将不同输入映射为高度相似甚至近乎相同的表示。在自监督和对比学习中，如果缺乏有效约束，模型可能找到低信息量解，形成完全坍缩或维度坍缩。Jing、LeCun 等系统刻画了坍缩现象，并解释 BYOL、SimSiam 等方法为何在不使用反例的情况下仍能避免坍缩。

避免坍缩的核心在于引入适当不对称和去冗余约束。典型方法包括使用反例、停止一侧梯度，引入动量编码器，或要求两路输出降低重复度。注意力头和专家网络之间若出现高度雷同，也可视为更广义的表征退化问题，需要通过架构和训练目标加以约束。

5.3. 表征的可辨识性

不同训练过程、数据来源和模态是否会收敛到相似表征，是可辨识性问题的核心。该问题关系到表征是否具有客观性，以及模型学到的表示能否对应真实世界因果结构。

柏拉图表征假说提供了表征趋同的一类证据。Huh 等提出，不同目标、数据和模态训练出的网络表征正在变得更相似，仿佛共同逼近现实世界的统计结构，且模型规模越大趋同性越强。CKA 等度量工具也可在不同初始化网络间识别对应层和对应表征。

因果表征学习进一步要求，好表征应还原数据背后的因果变量，而不只是拟合表面相关。Schölkopf 等系统提出因果表征学习框架，强调从底层观测中发现高层因果结构，以提升模型稳健性和泛化能力。

5.4. 什么是“好表征”

“好的表征”尚无统一定义。我们归纳出四类判据：1）任务准确率高，对应 CLIP、ImageBind 等评测逻辑；2）可解释、可迁移；3）与因果结构一致；4）能够解缠数据中的关键因素。

5.5. 多模态统一表征是否可能

多模态统一表征的核心问题，是图像、文字、声音和动作能否进入同一空间。CLIP 实现图文对齐，ImageBind 进一步将图像、文字、音频、深度、热成像、运动传感六种模态绑定到同一空间，并在跨模态检索和生成中获得迁移能力。柏拉图表征假说也从理

论层面支持不同模态表征趋同的可能性。

统一空间仍面临模态特征损失和模态鸿沟问题。我们认为，多模态统一表征已取得较大进展，但其是否真正形成共享空间、是否牺牲单一模态的细粒度信息，仍需后续任务和产业场景验证。

6. 物理 AI 落地：自动驾驶与机器人的表征分歧

物理 AI 将表征问题从数字空间推向真实世界交互。自动驾驶和机器人都需要将感知、状态、动作和环境变化编码为可决策的内部表示，差异在于数据来源、实时性要求和安全闭环约束。

6.1. 自动驾驶：端到端演进下的四类分歧

分歧一在于一段式与两段式架构。两段式先完成环境感知，再进行规划控制，中间结果如车道线、目标轨迹较容易排查；一段式则将感知与决策压缩到单一模型中，直接由传感器输入输出轨迹或控制信号，信息链路更短，但可解释性和故障定位难度更高。

分歧二在于纯端到端与 VLA 路线。VLA（视觉-语言-动作）在视觉输入与动作输出之间引入语言或语义中间表征，提升指令理解和解释能力。2025 年多家车企转向 VLA：理想发布“VLA 司机大模型”，输出轨迹和控制信号；小鹏第二代 VLA 去掉先转译为语言再行动的环节，改为视觉语言直接出动作；元戎启行将 VLA 称为端到端 2.0，强调可解释性。

分歧三在于世界模型的部署位置和使用方式。世界模型用于预测后续状态，一类用法是在云端生成罕见危险场景，支持训练和评测；另一类用法是在车端参与实时决策。华为 ADS 4 的 WEWA 架构将云端世界引擎与车端世界行为模型分工；蔚来 NWM 强调车端推演；特斯拉世界模型更多用于云端仿真和闭环评测。商汤开悟、Wayve GAIA 系列、英伟达 Cosmos 也主要服务于数据生成与评测。

分歧四在于显式生成与隐式潜空间。显式路线生成可见未来画面，隐式路线则在潜空间预测关键状态。自动驾驶量产侧显式生成式世界模型更常见，特斯拉、华为、蔚来、商汤、Wayve 均偏显式；隐式潜空间路线更多见于学术研究与机器人场景，车端量产阵营尚不清晰。

表5：自动驾驶主要公司/机构的表征路线

公司/机构	一句话说明路线	表征关键词
特斯拉	端到端 One Model；世界模型用于云端仿真与闭环评测	视频→控制；显式世界模型，云端，
华为	ADS 4 WEWA：云端世界引擎造数据 + 车端世界行为模型决策	显式世界模型双端分工 + MoE
理想	端到端 + VLM→VLA 司机大模型，强化学习，输出轨迹 token，	VLA，语言作中间表征
小鹏	第二代 VLA，去掉语言转译，视觉语言直接出动作	VLA，去语言中介，
蔚来	NWM 世界模型，100ms 推演 216 种场景，自带闭环仿真	显式生成式世界模型，车端推演，
商汤绝影	R-UniAD：强化学习端到端 + “开悟”世界模型协同	一段式端到端 + 显式世界模型
Wayve	GAIA 系列生成式世界模型，用于仿真与评估	显式生成式世界模型

数据来源：特斯拉、华为、理想、小鹏、蔚来、商汤绝影、Wayve 等公司官网，东吴证券研究所

6.2. 机器人：动作表征与数据来源是核心分歧

分歧一在于离散动作 token 与连续动作。RT-2 等路线将连续动作切分为离散动作 token，从而复用大语言模型处理序列的范式，优势是工程体系相对成熟，约束是动作输出可能不够平滑；π0 等路线用扩散或 flow matching 生成连续动作，支持 50Hz 等高频控制。机器人关节数量多、控制精度要求高，连续动作在高自由度任务中更具适配性。

分歧二在于训练数据来源。机器人真机数据稀缺且采集成本高，因此数据来源成为路线分水岭：1）从人类操作视频中预训练，如北大等 Being-H0 将人手视为基础操作器并学习毫米级动作 token；2）依赖真机遥操作数据，如 π0；3）依赖仿真与多源数据，如英伟达 GR00T N1。多源混合正成为重要趋势，智元 GO-1 同时使用跨本体、人类视频与百万真机数据。

分歧三在于单体 VLA 与快慢双系统。单体 VLA 试图由一个模型贯穿感知、理解和动作输出；快慢双系统则将场景理解与高频控制拆开，慢系统负责理解和规划，快系统负责连续动作生成。Figure Helix 采用慢系统理解与快系统 200Hz 动作输出，英伟达 GR00T N1 和星海图 G0 也采用类似思路；智元 GO-1 则在 VLA 中加入隐式动作中间层 ViLLA，形成理解、规划、执行三层分工。

分歧四在于显式生成与隐式潜空间世界模型。Meta V-JEPA 2 代表隐式潜空间路线，不生成画面、仅预测关键特征，并在部分零样本机器人控制任务中取得较高成功率；显式生成路线则包括 1X 物理世界模型、World Labs Marble 和 DeepMind Genie 2 等，重点是生成可观察、可交互的物理世界。李飞飞关于空间智能和世界模型的公开观点，也强化了显式世界建模的产业关注度。

表6：机器人和具身智能主要公司/机构的表征路线

公司/机构	一句话说明路线	表征关键词
RT-2 Google	把动作切成离散 token，复用语言模型范式	离散动作 token
π0 Physical Intelligence	flow matching 生成连续动作，50Hz，靠真机数据	连续动作，扩散/flow，
Being-H0，北大等，	从大规模人类视频预训练，人手为基础操作器	人类视频 + 毫米级动作 token
Figure Helix	双系统：VLM 慢思考 + 扩散快思考，120Hz	快慢双系统
英伟达 GR00T N1	开放人形基础模型，VLM + 扩散双系统，仿真训练	快慢双系统 + 仿真
智元 GO-1	ViLLA: 在 VLA 中加“隐式动作”中间层，三层分工	隐式动作 Latent Action
V-JEPA 2 Meta	隐式潜空间世界模型，零样本机器人控制	隐式潜空间 JEPA
1X / World Labs	显式生成式世界模型，看视频/造 3D 世界学任务	显式生成式世界模型

数据来源：Figure、英伟达、智元等公司官网，东吴证券研究所

6.3. 两个领域的共性与差异

共性在于，自动驾驶与机器人均围绕物理世界的表征形式展开路线分化。两者都需要回答动作或输出采用离散 token 还是连续数字，是否引入语言或多模态中间表征，是否采用一体端到端还是快慢分层架构，以及世界模型先用于云端数据生成和评测还是进一步参与端侧决策。

差异主要体现在三方面：1）数据来源不同，自动驾驶拥有较大规模车队真实里程，世界模型主要补足长尾危险场景；机器人真机数据稀缺，更依赖人类视频、仿真和遥操作组合；2）实时性与动作维度不同，人形机器人关节多、控制频率高，连续动作和扩散/flow 路线适配性更强，汽车输出维度相对较低；3）世界模型定位不同，自动驾驶更多将其作为造数据和评测工具，机器人则更期待其成为通用学习底座和空间智能载体。

7. 结论：表征革命是 AI 范式演进的主线

表征是理解 AI 范式演进的底层变量。它决定机器以何种方式组织外部世界，也是模型内部形成理解、生成和决策能力的基础。观察一家公司或一条技术路线，核心不只是看模型规模和参数数量，也要看其如何表征输入、状态、动作和未来变化。

每一次 AI 范式迁移，都伴随表征方式变化。从人工符号到专家特征，从端到端连续表征到跨模态基础模型，底层变化都是机器记录、压缩和组织世界方式的升级。当前

围绕表征中心与生成中心、符号与分布式、离散与连续、显式与隐式世界模型的分歧，说明新一轮表征范式仍处于验证阶段。

后续可沿五个方向观察：统一世界表征、离散与连续路线、显式与隐式世界模型、可解释可控表征，以及因果表征与真实世界数据。

方向一，统一世界表征能否成形。若图像、文字、声音和动作能够持续收敛至统一表征空间，通用智能体的底座将更加清晰。后续关注跨模态零样本能力是否持续提升、是否出现覆盖多模态多任务的标志性模型，以及模态鸿沟是否显著缩小。

方向二，离散与连续路线如何收敛。该问题决定下一代大模型和具身模型底层架构，是继续沿用 token 化范式，还是转向 JEPA 式连续潜空间。后续关注连续潜空间路线在主流基准上的稳定表现，以及头部实验室和产业团队是否持续增加相关投入。

方向三，显式世界模型与隐式潜空间谁主导物理 AI。该问题影响自动驾驶和机器人的技术路线与投资主线。后续关注世界模型是否真正进入车端或机端参与决策，而不只停留在云端数据生成和评测；同时关注隐式潜空间方案在真实任务中的稳健性是否追平或超过显式生成路线。

方向四，可解释与可控表征能否实用化。大模型在高价值和强监管场景落地，需要解决表征不可读、生成不可控和对齐不稳定等问题。后续关注表征探针、表征编辑、概念级可控生成等技术能否从论文走向产品，以及监管是否提出更明确的可解释性要求。

方向五，因果表征与真实世界数据。AI 要实现稳健泛化，需要表征更贴近真实世界因果结构；物理 AI 的瓶颈也正从算法本身延伸至高质量真实交互数据。后续关注因果表征学习能否在工业场景跑通，以及人类视频、真机和仿真数据能否形成可持续的数据闭环。

表征问题不会随具体路线变化而过时。如何更好地表征世界，是理解人工智能过去与未来的主线，也是判断技术路线和产业格局的重要抓手。

8. 风险提示

技术迭代不及预期的风险：数字 AI、物理 AI 技术进展低于预期，任务或应用场景拓宽速度低于预期。

大模型厂商 ARR 增速放缓的风险：若客户增长、续费率或客单价不及预期，模型厂商 ARR 增速可能放缓。

云服务商资本开支不及预期的风险：若自由现金流承压，或 AI 算力需求及投资回报低于预期，云服务商可能下调资本开支。

智能驾驶及机器人商业化落地不及预期的风险：若产品可靠性、成本下降或用户需求低于预期，相关产品商业化进程可能放缓。

免责声明

东吴证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本研究报告仅供东吴证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

在法律许可的情况下，东吴证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。

市场有风险，投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息，本公司力求但不保证这些信息的准确性和完整性，也不保证文中观点或陈述不会发生任何变更，在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。

东吴证券投资评级标准

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力相对基准表现的预期（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数，新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的），北交所基准指数为北证 50 指数），具体如下：

公司投资评级：

买入：预期未来 6 个月个股涨跌幅相对基准在 15%以上；

增持：预期未来 6 个月个股涨跌幅相对基准介于 5%与 15%之间；

中性：预期未来 6 个月个股涨跌幅相对基准介于-5%与 5%之间；

减持：预期未来 6 个月个股涨跌幅相对基准介于-15%与-5%之间；

卖出：预期未来 6 个月个股涨跌幅相对基准在-15%以下。

行业投资评级：

增持： 预期未来 6 个月内，行业指数相对强于基准 5%以上；

中性： 预期未来 6 个月内，行业指数相对基准-5%与 5%；

减持： 预期未来 6 个月内，行业指数相对弱于基准 5%以上。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议。投资者买入或者卖出证券的决定应当充分考虑自身特定状况，如具体投资目的、财务状况以及特定需求等，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。

东吴证券研究所
苏州工业园区星阳街 5 号
邮政编码：215021
传真：（0512）62938527
公司网址： http://www.dwzq.com.cn