

## 人工智能国内外跟帖(1~645 条)评论大纲

- **评人工智能向何处去？**
- **机器学习/深度学习/增强学习（不可解释性）**  
带来今天人工智能应用场景的繁荣（不可解释的机器学习）
- **可解释的机器学习**  
成为今天人工智能发展的亮点
- **基于异步脉冲神经网络的神经拟态计算系统**  
已见到突破的曙光
- **依托大规模语义网络（知识图谱）支持，实现认知智能解决方案**  
还差最后一公里
- **脑机接口的理论和实践**  
国内外已有十几个案例
- **探索下一代通用人工智能（强或超强人工智能）**  
（提出概念，质疑人工智能各不同学派提出的发展途径）
- **对下一代通用人工智能的质疑**

# 评人工智能向何处去？

（兼谈国内外跟帖评论）

陆首群 2021. 1. 15

2019年8月8日，针对当时一个世界性的热门话题：人工智能如何走向新阶段？基础理论抓什么？我们感到有必要建立一个公开评论的平台（在CSDN网站上发布），邀请或吸引国内外专家、草根（不拘一格）以跟帖方式参与讨论集思广益。到2021年1月15日（约一年半时间），我们公布了来自国内外的645条跟帖，其中不乏真知灼见，并且完全覆盖今天国内外人工智能发展前沿！我们创办跟帖讨论平台有一个优势：我们在国内外科技界、企业界高层有广泛的人脉，几十年来我们在国内外结识了不少开源的、ICT的、AI的以及数学的、哲学的专家朋友（其中不少大师），以及朋友的朋友；人工智能讨论平台是以国内外专家评论（包括摘引的）为骨架，以发布跟帖方式启动的。

众所周知，中方发展人工智能受到美国打压。如果中外专家名人要迈过跨洋跟帖就人工智能交换意见，很可能被纳入黑名单。以致难以成事！这时我们发表公开声明：我们倡导的跨洋跟帖属于公开的、开源的性质，以此对抗打压，受到国内外专家欢迎。公开的、开源的提法具有免疫力！

在已发表的645中，从评论机器学习/深度学习那些不可解释的人工智能的初级阶段开始，而今天人工智能的繁荣正是基于机器学习/深度学习，说它们已近天花板（已经没有发展潜力）是不妥的。谈到人工智能的出路在何处？归纳大量跟帖中列出的4条（包括研发基础理论）：①打破机器学习的黑盒子研发可解释的人工智能，②基于异步脉冲神经网络的神经拟态计算系统，③依托大规模语义网络（知识图谱）的支持，实现认

知智能解决方案，④脑机接口的理论和实践。目前①已成为全球的热点，②已见亮点（国内外已有先例），③还差最后一公里，④国内外已有十几例试点（用于帕金森病、中年忧郁症、儿童自闭症的诊治及中风、癫痫病人的辅助治疗）。

最近，有关专家质疑由人工智能不同学派提出的上述四个发展途径，他们认为人工智能不同学派互不相容，单打独斗，缺乏哲学的和科学的全面统一范式的理论基础，缺乏整体观，他们只是从不同侧面模拟人类心智（大脑），各自提出的“发展路径”均有片面性。他们的建议是从改革、融合、统一的人工智能发展范式出发，提出发展下一代强（或超强）人工智能的思路。他们也将其思路汇集到跟帖中来讨论。可是他们提出的思路尚属于概念或构想，离解决方案还差得很远，因此他们提出的质疑刚一露头，就有人提出质疑的质疑。我们欢迎这场辩论。

必须指出，当今世界已进入“量子计算+人工智能+基因科学”新时代，而且时代之声呼吁拥抱开源。在发布人工智能跟帖时，也将发布与之密切相关的量子计算（量子统信）、深度信息技术、基因医疗的跟帖。

现在的问题是针对“人工智能向何处去？”的跟帖评论要不要继续办下去？回答是：深受国内外欢迎，呼呼应继续办下去！本集是人工智能国内外跟帖评论的第八集（585条~645条），我们还将办好“人工智能国内外跟帖评论续集”，从第九集（或从646条跟帖）开始！

## 国内外 AI 跟帖留言 ( 585 条-645 条跟帖 )

**585**, 基于时间序列任务深度学习模型的可解释人工智能技术实证研究

(An empirical study explain able AI techniques on deep learning models for time series tasks)

—— UdoSchlegel , DanielaOelke , DanielA-keim , MennatallahEl-Assady

康斯坦茨大学, 奥芬堡应用科技大学

机器学习黑盒子模型的决策解释通常是通过应用可解释 AI (XAI) 技术生成的。但是许多建议的 XAI 方法的结果通常是未验证白输出。通常是通过人类手工对单个图像或文本进行视觉解释来实现对模型解释的评估和验证。本文作者提出了一项经验研究理论和基准框架, 用来将归因方法应用于为时间序列上的图像和文本数据开发的神经网络。这种方法可以使用扰动方法来自动确定时间序列的归因技术并识别其可靠的方法。

可查阅 <https://arxiv.org/pdf/2012.04344.pdf>

2020.12.08 发表

**586**, 可解释人工智能: 训练数据的子集如何影响预测

Explainable Artificial intelligence: How subsets of the Training Data Affect a prediction——AndreasBrandsaeter, IngridK.Glad

2020.12.7 发表

(可查阅 <https://Arxiv.org/pdf/2012.03625.pdf>)

各种应用领域中, 对机器学习模型和预测的解释与说明的兴趣及需求日益增长。本文考虑已经被开发、实施和训练的数据驱动模型, 其目标是解释模型并解释和理解其预测。

由于数据驱动模型所做的预测严重依赖于用于训练的数据, 因此作者认为解释应传达有关训练数据如何影响预测的信息, 为此提出一种称为 shapley 值, 用于衡量训练数据子集的重要性。shapley 价值概念源自博弈理论, 其发展目的是一组合作的参与者之间公平地分配支出。

作者描述和说明所提的方法是有用的，并在几个示例上证明其功能。从而展示了如何将提出的解释用以揭示模型和错误的训练数据中的偏差。

## 587, 英特尔公布神经拟态新进展, 探讨神经拟态计算的商业应用, 更新 Loihi 神经拟态研究基准

——浙江大学融媒体中心谢子怡报导, 2020. 12. 5

在 12 月 4 日英特尔研究院开放日上, 英特尔神经拟态计算应用合作伙伴: 联想、埃森哲、罗技、梅赛德斯-奔驰、机器视觉传感器公司 PropheSee 等与英特尔一起分享英特尔神经拟态研究社区 (INRC) 的最新应用成果 (进展) 及其在商业应用上的价值。这次英特尔重点介绍 Loihi 神经拟态研究基准, 包括:

### ■ 语音命令识别

埃森哲测试了英特尔 Loihi 和标准 GPU 上识别语音命令的能力, 发现 Loihi 不仅达到了和 GPU 类似的精度, 而且能效提高 1000 倍以上, 响应速度快 200 毫秒。梅塞德斯-奔驰也正在探索如何将这些结果应用到现实中, 如在汽车中加入新的语音交互命令。

### ■ 手势识别

埃森哲展示了 Loihi 在快速学习、识别个性化手势方面取得的切实进展。只需几次曝光, Loihi 即可学习新手势, 可用于智能产品交互, 或者公共场所非接触显示。

### ■ 图像检索

零售行业的研究人员评估了 Loihi 对基于图像的产品搜索应用, 发现在保持相同精度水平的情况下, Loihi 生成图像特征向量的效率, 比传统 CPU、GPU 方案提升 3 倍多。英特尔此前研究发现, Loihi 在百万幅图像数据库中搜索特征向量的速度比 CPU 快 24 倍, 且能耗低 30 倍。

### ■ 优化和搜索

英特尔发现 Loihi 解决优化和搜索问题的效率比传统 CPU 提高 1000 倍, 速度快 100 倍, 此研究可用于无人机实时规划并做出复杂导航决策, 也可以扩展到复杂的数据中心负载, 完成协助列车调度、物流优化等任务。

### ■ 机器人技术

罗格斯大学和代尔夫特理工大学的研究人员展示了在 Loihi 上运行机器人导航、微型无人机控制应用。代尔夫特理工大学的无人机使用一个包含 35 个神经元且能演进的脉冲网络进行光流着陆 (Opticflowlanding)，频率超过 250HZ。罗格斯大学发现，在同等性能下，Loihi 解决方案的功耗比传统 GPU 低 75 倍。Loihi 还可以成功学习诸多 OpenAIGym 的任务，精度与深度行动者网络 (DeepActorNetwork) 旗鼓相当，而能耗比移动 GPU 解决方案低 140 倍。英特尔还展示了 Loihi 如何自适应地控制水平跟踪无人机平台，实现最高 20KHZ 的闭环速度、200 微秒的视觉处理延迟比传统方案提高 1000 倍。

### 588, 用于训练脉冲神经网络的多代理进化机器人框架

(A multi-agent evolutionary robotics framework to train spiking neural networks)

普渡大学 SouvikDas, AnirudhShankar, VaneetAggarwal

2020.12 发表

(可查向阅 <https://Arxiv.org/pdf/2012.03485.pdf>)

自然界中广泛存在着竞争进化的机制。受此启发，作者提出一种基于多代理进化机器人 (ER) 的新框架用于训练脉冲神经网络。脉冲神经单元的权重和它们控制的机器人的形态参数被视为表型，框架规则根据某些机器人和它们的脉冲神经单元在生存和繁殖过程没有明确的变异，但它们通过狩猎食物以及在规则中存活来隐式驱动进化。它们捕获食物的效率在训练中表现出点缀平衡的进化特征。作者演示了表型上的两种进化继承算法——突变和突变的交叉，比单纯突变有着显著优势，提升了 40% 的学习能力。

### 589, 将上下文融合到常识推理的知识图谱中 (Fusing context into knowledge graph for common sense reasoning)

——YichongXu, ChenguangZhu, RuochenXu, YangLiu, MichaelZeng, XuedongHuang

微软认知服务研究组

(可查阅 <https://arxiv.org/abs/2012.04808>)

人工智能的一个关键方面是基于观察和知识对日常事务进行推理的能力。通常，大多数人都将这种功能作为与世界交流和互动的主要基础。因此常识推理成为自然语言理解中的重要任务。在这一领域已提出了各种数据集和模型。尽管诸如 BERT 和 RoBERTa 之类的大规模预训练模型在语言理解方面很有效，但它们缺乏明确处理知识和常识的模块。而且与结构化数据相比，本文在表示常识方面的效率要低得多。因此存在各种方法将语言模型与各种形式的知识图耦合以进行常识推理，包括知识库、关系路径、图关系网络和异构图。这些方法结合了语言建模和结构知识信息的优点，并提高了常识推理的性能。但是这些模型与人类的智能之间仍在不可忽略的差距。一个原因是，尽管知识图谱可以在概念之间编码托朴信息，但它缺乏丰富的上下文信息。另一方面我们可以从外部来源获得精确定义。因此要生成可以无缝集成到语言模型中的结构化数据的表示形式，我们需要在知识图谱中提供每个概念的全景视图，包括其相邻概念，与之的关系以及对其的明确描述。

因此作者提出了 DEKCOR 模型，即用于常识推理的描述性知识。给定一个常识性问题和选择，首先提取所包含的概念。然后在 ConceptNet 中提取问题概念和选择概念之间的边界。如果不存在这样的边缘，作者将为每个包含选择概念的三元组（节点、边缘节点）计算相关性得分，并选择最高的得分。接下来通过各种文本即匹配条件从维基词典中检索这些概念的定义。最后，将问题选择的三元组和定义输入语言模型 Albert 中，并且相关性分数由附加的注意层和 Soft max 层生成。作者在 Common sense QA 数据集上评估了我们的模型，并且 DEKCOR 在测试集中的表现优于之前的最新结果 1.2%（单个模型）和 3.8%（整体模型）成为第一个超过 80% 的准确模型。

**590**，一个基于显性和隐性交互的广义相加模型的可解释性推荐系统

——YifengGuo, YuSu, ZebinYang, AijunZhang

香港大学保险学院统计系，深圳索信达控股公司

（可查阅 <https://Arxiv.org/abs/2012.08196>）

在过去的几十年里，预测用户喜好的推荐系统被广泛应用于多个领域，如电子商务、社交媒体、银行等。本文从统计建模的角度，基于显性交互和

隐性交互的广义相加模型 (Generalized Additive Model, GAM)，提出一个可解释性推荐系统 GAMMLI。

该系统可捕捉观察到特征的主要影响和显性交互，并发现未观测到特征的隐性交互，所有影响都可以用可视化的方式来解释。

与传统协同的过滤方法不同，GAMMLI 考虑了用户和项目的群体效应，这有利于提高模型的可解释性，也有利于冷启动推荐问题 (Cold - start recommendation problem)。实验结果表明，GAMMLI 在预测性能和可解释性方面都具有优势。

(注：冷启动问题，如何在没有大量用户数据的情况下设计个性化推荐系统并让用户对推荐结果满意从而愿意使用推荐系统)

### 591, XAI-P-I: 从实践到理论的可解释人工智能

(XAI-P-I: A Brief Review of Explainable Artificial Intelligence view of Explainable artificial intelligence from Practice to Theory)

——NazaninFouladgar, KaryFramling

2020.12.17 发表

(可查阅 <https://arxiv.org/pdf/2012.09636.pdf>)

随着新模型的出现，机器学习已经在不同应用中得到了巨大的应用，并随着时间的推移，机器学习也在不断地发展。然而这些模型的复杂行为阻碍了人们简洁地理解如何做出特定的决策。此限制要求机器通过解释来提供透明性。因此可以将“黑盒子”分配给机器学习模型以做出决策，而将“白盒子”分配给这些模型的解释版本以“解释 AI”的主题。

问题表明，解释可以以因果关系和非因果关系的形式来说明。尽管因果关系已在 AI 研究人员中解决算法决策问题变得更加普及，但非因果关系最近吸引了人机交互领域的学者。实际上，XAI 尚未成熟，并且有很多开放的空间可以提传实践和理论上的解释。尽管解决这些问题十分关键，从实践和理论上仍然缺乏对 XAI 领域目前的现状作简明的理解。

本文首先关注黑盒子解释的类别并给出一个实际的例子，然后讨论如何以多学科为基础讨论理论解释，最后提出未来工作的一些方向。



## 592, 用通过贝叶斯学习的二元加权法来训练脉冲神经网络

( BiSNN : TrainingSpikingNeuralNetworksWithBinaryWeights ViaBayesianLearning)

——HyeryungJang, NicolasSkatchkovsky, OsvaldoSimeone

伦敦国王学院, 2020.12.15

(可查阅: <https://arxiv.org/pdf/2012.08300Vladimir.pdf>)

通过将突触权重限制为二进制, 可以采用电池供电设备的基于人工神经网络(ANN)的计算更加节能, 从而无需执行乘法运算。另一种新兴的方法依赖于使用脉冲神经网络(SNN), 这种生物启发、动态的、事件驱动的模式, 通过使用二元、稀疏、激活的方法来提高能源利用效率。

本文介绍一种脉冲神经网络模型, 该模型通过引入一个具有二进制突触权重的SNN, 由此产生的二进制SNN模型(称BiSNN)不仅能够通过使用尖峰神经元来利用时间稀疏性, 还可以通过二元加权降低神经操作的复杂性。通过在合成和真实神经形态数据集上的实验, 验证了关于全精度实现的性能损失, 并展示了贝叶斯范式在准确性和校准方面的优势。得出两个学习规则, 第一个基于Straight-through和Surrogate梯度技术的结合, 第二个基于贝叶斯范式。实验验证了相对于全精度实现的性能损失, 并证明了贝叶斯范式在准确性和校准方面的优势。

## 593, 研发脑机接口技术使难治的抑郁症治疗获新机

2020年12月11日上海交通大学医学院附属瑞金医院脑机接口及神经调控中心正式成立, 第一个临床脑机接口研究项目——“难治性抑郁症脑机接口神经调控治疗临床研究”同日正式启动。全球至少有3.5亿受抑郁症困扰的患者, 将借助最前沿的脑机接口技术来治疗。

脑机接口应用前景广泛, 如: 癫痫或帕金森病、抑郁症、精神分裂症、强迫症、厌食症、成瘾疾病如酒精依赖等。

## 594, 在脑机接口技术研发中基于脑功能网络和样本熵做脑电信号特征提取

——杭州电子科技大学智能控制与机器人研究所罗志增、鲁先举、周莹

2020.12.17发表

针对脑机接口（BCI）研究中采用单一特征对运动想像脑电信号（EEG）识别率不高的问题，本文提出了一种结合脑功能网络和样本熵的特征提取方法，以提高脑电信号识别率。

**595，** 日产公司发布设计 2050 年的概念车，利用脑机接口技术设计

日产于 2020 年 12 月 16 日对外发布一款名为 GT-RX2050 的全新概念车，该车是日产美国设计公司（NDA）的 JaebumCho 设计。他对 2050 年的车型进行了大胆的畅想，并基于日产跑车 GTR 的灵感，依托脑机接口技术，设计制造这款概念车。

所谓依托脑机接口技术，即通过大脑信号直接控制计算机，进而控制车辆。设计师表示，如此设计在于创造一种外骨骼，使人车合一。这款概念车设计的本质是构想一种新的交通方式（不是传统的交通工具）。

据了解这款车现已制造出 1:1 的概念车。

**596，** 邁向可持续的千倍速计算未来

——英特尔中国研究院院长宋继强

回顾 2020 年，新冠肺炎疫情给我们的工作和生活模式带来了重大变化。其中，有一项技术正在潜移默化地改变我们的生活，即 5G。2020 年是 5G 商用的关键之年，中国的 5G 技术部署在这一年取得了突飞猛进的进展。5G 是由技术驱动的创新，早在约 10 年之前，5G 技术就开始投入研发。之所以要大力拓展 5G 技术，并不是为了迎合当时的需求，而是看到了未来对于带宽和网速的需求潜力，是为“未来”做出的技术布局。

5G 技术的进展让我想到了近期英特尔研究院开放日的活动主题——“追求计算的千倍提升”。类似于 5G，要满足未来的计算需求，即超高带宽、超低时延、超大规模连接的需求，我们需要一种“超前”思维。因此产业现在就要开始提前布局，追求计算的千倍提升，在目前智能化、数字化的大背景下，这种“超前”思维非常有必要。

**以超前思维布局未来计算范式**

数字化、智能化已经成为不可阻挡的趋势，受到今年新冠肺炎疫情的影响，这一趋势以更快的速度席卷而来。如今，全球已有超过 100 亿台设备与云中的超级计算机实现了互联，未来这一数字将增长到 1000 亿。拥抱数字化不是选择题，而是必选题。在全民数字化的浪潮之下，数据量呈爆发式增长态势，数据形式也更加多元化，可以说，未来的计算需求将有千倍速的提升。英特尔追求计算的千倍提升，就是从计算的供给侧出发，为未来的计算需求构建坚实基础。

除了“超前”思维之外，要想实现计算的千倍提升，还需要“超常”思维，即要打破常规。随着数据越来越多元化，新的计算范式不再是锦上添花，而是雪中送炭。常规和传统的单一架构已经不能满足越来越复杂的计算需求，未来需要更快、更灵活、更低功耗的“新计算”来破题。

### 软硬件双突破释放千倍算力

这种“超常”思维将在以下几个领域得到体现。首先，在硬件方面，需要打破单一架构，多架构融合的 XPU 架构将成为主流。XPU 架构不仅能大幅提升算力，同时还能够根据需求进行快速组合，降低成本，灵活性高。英特尔是目前全球唯一一家已经覆盖这四种主流芯片的厂商，得益于先进的封装技术，英特尔正在异构计算领域突飞猛进。

除此之外，面向未来，还需要对架构本身践行“超常”思维。举例来说，颠覆传统的冯·诺伊曼架构，模仿人脑神经元结构的神经拟态计算芯片就是一个很好的例证。这种芯片的优势在于可以在提升性能的同时大幅降低能耗。英特尔及其合作伙伴发现，英特尔神经拟态计算芯片 Loihi 解决优化和搜索问题的能效比传统 CPU 高 1000 倍，速度快 100 倍，已经实现计算速度的千倍提升。

其次，在软件方面。XPU 架构的诞生，给软件提出了更高的要求，因为能够同时掌握多种架构编程语言的开发人员凤毛麟角，而软件是释放硬件性能的关键一环，因此能够跨架构编程的软件模型以及可以提升编程效率的

工具就显得极为重要。为此，英特尔也提前布局，跨架构编程的统一模型 oneAPI Glod 版本已在本月正式发布，将在很大程度上解决跨架构编程的难题。

## 未来算力要强大也要绿色

要实现算力的千倍提升，还需要坚持可持续发展的原则。千倍速的提升不能以千倍的功耗为代价，可持续发展是实现千倍提升的必要条件。

目前，计算对于能源的需求巨大。有研究报告显示，训练一个大型 AI 模型所产生的碳排放量相当于 5 辆美式轿车整个生命周期所消耗的碳排放量。因此面向未来计算的千倍提升，只有坚持可持续发展原则，才是真正符合人类利益的技术进步。

英特尔在技术发展方面一直坚持可持续原则。已经有结果显示，作为下一代 AI 芯片，英特尔神经拟态计算芯片 Loihi 在处理语音命令识别时，不仅达到了和 GPU 类似的精度，并且能效提高 1000 倍以上。除此之外，英特尔最新的集成光电技术将光子技术与硅芯片紧密集成，可以最大限度地缩小硅光子设备的体积，从而降低成本，将对数据中心进行彻底革新。

诸如此类的例子在英特尔还有很多，英特尔的宏旨是“创造改变世界的科技，造福地球上的每一个人”，通过我们的“超前”思维、“超常”思维以及可持续发展的原则，英特尔正引领产业迈向千倍速的计算未来。对这一天的到来，我充满期待。

### 597, 10 例脑机接口案例

跟帖 12 条，美国卡耐基梅隆大学，2019.8 发布

跟帖 51 条，俄罗斯 Neurobotics 公司和莫斯科 MIPT 学院 2019 发布

跟帖 52 条，美国脸书和加州大学 UCSF 分校 2019.7 发布

跟帖 175 条，美国特斯拉 2019 宣布

跟帖 308，中国浙江大学 2020.1 发布

跟帖 363, 美国加州大学伯克利分校 2019 发布

跟帖 378, 美国加州大学旧舍山分校 2019 发布

跟帖 386, 中国天津大学 2020.1 发布

跟帖 386, 中国启动天宫二号, 2020.1 发布 (不用开颅在脑中嵌埋芯片采用脑外感应测试)

跟帖 593, 上海交大附属瑞金医院 2020.12.11 发布

**598, 探索构建符合中国国情的人工智能治理框架**

——汪洋主席在十三届全国政协第 45 次双周协商座谈会上讲话摘要, 2020.12.23

任何一项重大科技创新都会促进社会生产力和生产关系的变化, 也会影响社会生活方式的伦理秩序的调整。人工智能技术具有广泛渗透性和颠覆性, 带来的科技伦理与法律问题更加复杂多元, 可能引发潜在伦理和法律风险。

**599, 人工智能安全治理**

——傅莹在首届清华大学人工智能合作与治理国际论坛上讲话摘要, 2000.12.18

在人工智能治理方面, 人类必须汲取历史教训, 避免像核武器、互联网等由于治理共识形成太晚, 而导致人类曾面临巨大威胁。

在人工智能安全治理方面, 既然人工智能武器化不可避免, 专家们的研究方向是寻找合适的治理路径。希望对人工智能, 尤其是智能武器的治理, 人类应走在技术变革的前面, 充分认识其风险, 早些达成治理共识。

**600, 利用深度学习快速准确识别月球上的陨石坑**

中国探月工程团队在《nature》上发文: 利用深度神经网络训练以及迁移学习策略快速准确识别月球撞未坑, 新识别出的陨石坑有 10 万多个, 同时建立了月球陨石坑数据库。

**601, 量子拟态计算 (Quantumneurmorphic computing)**

——2020.10.13

(可查阅 <https://aip.scitation.org/doi/10.1063/5.0020014>)

量子拟态计算（注：拟态计算即类脑计算）通过量子器件在大脑机制启发下物理地实现神经网络从而加速计算。本文证明了该案例可用于中型量子计算机。讨论了数字和模拟电路的量子拟态神经网络不同的实现方式。

## 602, 人工智能面对的一些挑战

——姚期智（图灵奖获得者、中科院院士、清华大学交叉信息研究院院长）在《2020 浦江创新论坛》上的讲演

人工智能有三大技术瓶颈亟待突破，需要科学家“从 0 到 1”的原创研究。脆弱性和不稳定性是人工智能面临的第一大技术瓶颈。

人眼识别十分稳定，一个图像如有微小改变，人仍能一眼看出它是什么，然而人工智能在图像识别方面有点“人工智障”，比如将一只小猪的照片加入一些图像“杂音”，一些机器视觉系统居然会把它识别为飞机。“小猪变飞机”这种漏洞给人工智能应用带来安全隐患，比如黑客可以攻击汽车自动驾驶系统，诱导它将马路上的“停止”标识当作“通行”，从而引发交通事故。

第二大技术瓶颈是机器学习算法缺乏可解释性，很多算法处于“黑盒子”状态。例如一个科研团队开发了一个房地产估价系统，这个系统通过一套算法学习了有关各地房地产价格的大数据，从而能自动评估房地产价格。然而，这套算法像黑盒子一样，很难给出估价的完整依据。这在商业应用上是一块很大的短板，房地产商会怀疑：人工智能系统是否低估了价格？是不是有利益相关方对系统做了手脚，故意压价？因此，算法的可解释性问题亟待科研突破，否则会大幅限制人工智能的商业应用的进程。

第三大技术瓶颈是人工智能的对抗性较弱。

如今，一个无人机群可以轻松完成灯光秀、农林作业等任务，但要看到，这些任务都是在自然环境下完成的，如果是处在高对抗的人为环境中呢？比如在电子竞技和军事战斗中，无人机群的协同作战能力的强化学习、博弈论研究，让无人机群能够在高对抗环境中自主找到最优策略。

## 603, 可解释的抽象训练数据集

(ExplainableAbstractTrainsDataset)

葡萄牙里斯本 NOVA 大学

ManueldesousaRibeiro, LudwigKrippahl

2020.12.15 发表

(可查阅 <https://arxiv.org/pdf/2012.12115.pdf>)

可解释的抽象训练数据集是一个包含训练简化表示的图像数据集。它旨在为证明和解释提取算法的应用和研究提供一个平台。该数据集随附一个本体，该本体基于它们的视觉特征对所描绘的训练进行概念化和分类，从而可以精确地了解每个训练的标记方式。数据集中的每个图像都用描述训练特征的多个属性和训练元素的边框标注。

#### 604, 在记忆性脉冲神经网络中学习 STDP 模式识别的必要条件

(Necessary Conditions for STDP-based Pattern Recognition Learning in Memristive Spiking Neural Network)

Demin V. A. 等, 莫斯科大学和莫斯科国家研究中心

发表于《Neural Network》2020.12

本文目的在于研究实验和理论方法，通过高性能的基于记忆器的脉冲神经网络 (SNN) 寻找有效的本地训练规则，实现无监督模式识别。首先，使用脉冲——定时依赖性 (STDP) 进行权重变化的可能性研究，通过 CoFeB 连接的一对硬件模拟神经元来演示纳米合成记忆器。接下来，在单层完全连接的馈送 SNN 的多种记忆 STDP 参数中，对二元聚类任务解决方案的学习收敛进行了分析。显示出神经元之间具有优胜者——接受——全部竞争。为了研究训练收敛的基本条件，作者构建了一个基于速率的单层网络的原始概率生成模型，该模型具有独立或相互竞争的神经元，并进行了深入分析。主要结果是“相关性增长——反相关衰减”原则，它说明要接近最佳策略来配置模型参数。可定义为最佳学习的必要条件。最后，本文对一种启发式算法进行了实验，找出记忆 SNN 中的收敛条件。该方法具有通用性，在搜索本地规则以确保其在模式识别任务域中的无监督学习收敛时，可应用于软件或硬件速率编码单层 SNN 的多种电子学和神经元。

#### 605, 使用高成功率和高精度的量子加速脉冲神经网络

——北京大学、中国邮电、日本 KDDI 研究所

本文提出了一种量子脉冲神经网络，对其进行了综合评估，并给出了详细的数学证明，包括成功概率、计算精度和算法复杂度。证明表明量子 SNN 的计算复杂度在数据维度上是对数多项式。此外，本文还提供了一种将量子 SNN 的最小成功概率提高到接近 100% 的方法。最后本文展示了量子 SNN 在解决现实世界中模式识别方面的良好性能。

#### 606, 脑电与计算机视觉的交叉

——普渡大学电子计算机工程学院，2020.12.23 发表

发表在人工智能顶刊（TPAMI）上的一篇文章表示脑电与计算机视觉的交叉存在缺陷和盲点。

利用 EEG 测量技术，对观看 ImageNet 刺激的受试者的大脑诱发进行分类，并采用从该处理中获得的表示来构建一种新型的目标分类器是当下热门的研究方向。

研究结果说明了存在于所有神经影像数据中的时间自相关对分类实验有深远影响。

#### 607, 基于宽场成像（Wide-field imaging）的脑机接口

Kelly Clancy of Sainsbury Well come Centre

在《Neuron》杂志上发表（2020.12.27）

专家能够在人类或动物使用 BMI 的同时查看大脑皮层的整个背面。

这项技术使皮质的无偏屏幕能够定位与学习有意控制光标有关的区域。

#### 608, IBM 向 LFAI 捐赠 AIX360 以助力可解释 AI 实践

——IBM 田忠博士

早期的 AI 实践往往具有自解释的特点，因为那时使用的是规则库、决策树、抉择表等这类比较直观的技术。目前机器学习/深度学习日益普及，但这类技术的模型对于用户而言往往是黑盒子，需要所谓“事后分析解释”（Post-hoc-interpretation）来帮助用户打破黑盒子，建立对于该系统决策的合理信心。



“事后分析解释”既可以用来理解所使用的数据，也可以用来理解所训练出的模型。对于前者，可以采用 DIP-VAE 算法以提取最有效特征，可以采用案例式推理算法 ProtoDash 建立典型案例。对于后者，可分为全局解释和局部解释。全局解释指的是向用户展示该系统的整体预期决策模型，以帮助用户理解系统决策的合理性，局部解释就特定案例进行分析，找出影响模型做出该结论的关键要素。

目前，可解释 AI 的发展能够就特定问题向用户呈现有意义的解释。AIX360 是对于这些训练数据和模型建事后分析解释的工具包。研究人员给出了一个银行信贷决策系统的可解释实践。该 AI 系统基于全美公井真实数据 (FICOHELOC Dataset) 来辅助决策是否批准一项贷款申请。项目建设团队需要在验收阶段向银行主管解释本系统的决策效果，也就是要对所训练出的模型做出全局性的直观解释。因此，他们选择了 BRCG 和 GLRM 相互补充的规则生成算法，他们从当前的训练数据集上建立真值表，从而建立如下规则：

“没有至少 5 个户头或者拥有多过 5 个户头但负债超过 1000 美元的客户是高风险客户”

信贷经理关心的是这个系统给出的决策建议是否有系统歧视。他们用手边的正反案例来研究系统决策的合理性。ProtoDash 算法可以满足要求。经验算信贷经理发现：能获得贷款的申请者过半数的特征值都是接近的，不能获得贷款的申请者其半数特征也是接近的。这给了他很大信心。同时他还发现未来能获得贷款的申请者大多有轻微违法记录，这可能有过于严苛并有失公允，要额外小心处理。

## 609, 深度学习中的理论问题

(这些理论有助于人们对人工智能这个黑盒子的理解)

MIT 研究团队 Tomaso Poggio, Andrzej Banburski, Qianli Liao

(可查阅 <https://www.pnas.org/content/117/48/30039>)

虽然深度学习在许多应用程序中都是成功的，但理论上尚不十分了解。深度学习的理论表征应回答有关近似/优化动力学和样本外性能。

本文阐述在近似方面做出的工作，对深层/浅层架构近似能力的对比，训练神经网络的优化过程，尤其是梯度流对应动力学系统的属性，对复杂性控制方面也进行了理论表述，如对于特定类型的复合函数，卷积深度网络可以避免维数灾难。这些理论有助于人们对于人工智能这个黑盒子的玩理解。

## 610, 语义和解释：为什么反事实的解释会在深度神经网络中产生对抗性示例

( Semantics and explanation: Why counterfactual explanations produce Adversarial examples in deep neural network s)

Kieran Browne, Ben Swift research school of Humanities & the Arts  
Australian National University

如果不首先解决语义的稀缺性，将无法解释深度神经网络。已经存在计算方法来产生模型不可知的解释。当将这些方法应用于 DNN 常见的模棱两可或低级表示时，它们根本不能作为解释。这不仅仅是对现有解释方法的限制，而是没有语义就不可能有任何解释。由于深度学习通常对“原始数据”进行操作，几乎没有语义内容（例如像素和字符），因此这种实现有助于阐明可解释性的挑战，我们要么找到一种方法来提取假定存在于网络的隐藏层中的语义，要么承认失败。

最近在可解释人工智能（XAI）中的论文已经确定了该领域的理论基础存在的问题。该领域通常仅凭直觉来解释什么是解释，与人类解释和理解解释的方式不同，有人建议 XAI 根据心理学和社会科学研究的一系列原则采用“日常解释”。有的学者另外指定一种产生反事实解释的方法。认为反事实的解释、表演，与模型无关，可自动计算，并且对外行人员易于理解。作者认为，这些反事实的解释为向任何人解释复杂的算法系统提供了途径。但是，自 2014 年以来，等效计算已用于深度学习研究中，尽管没有给出解释。相反，在深度学习研究的背景下，反事实计算会产生“对抗性示例”，潜移默化地修改了输入，导致网络莫名其妙地分类错误。

这应该让我们思考：同一方法怎么可能一方面代表一种有前途的新方法向任何人解释深度神经网络的决策，另一方面又代表同一决策过程中令人困

惑的脆弱性？我们称这种现象为解释性分歧。作者认为这种鸿沟揭示了XAI研究在语义上的盲点。

## 611，基于人工智能技术设计的第六代战斗机纷纷亮相

（这里的人工智能技术是不可解释的机器学习/深度学习）

◎重录 283 条跟贴：

由人工智能控制的全球第一架六代机亮相

据《强武精兵》网站 2020.2 报导，全球第一架六代机——英国“暴风雨”亮相

目前在五代机尚未全部普及之时，全球六代机的竞争已经悄无声息的展开了，作为当前人类科技和最高智慧的结晶，全球第一架六代机亮相了，超过目前所有美俄战斗机。

所谓六代战斗机指由人工智能控制的吸气式高超音速战斗机。

◎去年（2020）年中获悉，美国全球第一架六代机试航首飞成功。该机采用人工智能技术设计（估计是不可解释的机器学习）。

◎据《兵工技术》2020.12.3 报导，中国研制六代机，研发自适应变循环发动机、综合航电系统（采用光传控制核心技术）、全相全频谱隐身技术（及纳米隐身材料）、太赫兹雷达、高超音速等新技术，估计也采用不可解释的机器学习/深度学习的人工智能技术。

## 612，大规模语义网络支持实现认知智能

从跟贴 186、190、210、217、234、235、236、237、238、239、245、249、263、265、277……中，谈到人工智能研究的难点是如何解决对认知的解释与建构，或者说，依托大规模语义网络的支持，以破解认知智能的解决方案，还差最后一公里。

大规模语义网络或知识图谱有希望成为大脑（即类脑），它是实现认知智能的关键技术，但目前大规模语义网络存在的缺陷：①自然语言从大脑思维的模拟讯号向机器能够识别/理解的数字讯号转变的精准度不足，难以推动计算机建模或难以实现认知智能，②目前大规模语义网络最大的问题是缺乏常识（及缺乏行业知识、专家经验和语言学知识等），或者说机器

最大的问题是缺乏常识。这样，大规模语义网络尚无力支持智能体提高其对认知的理解和解释能力，还差最后一公里。

应该指出，2019 年大规模语义网络取得了飞跃式发展，提高了语言转变（如从模拟讯号转变为数字讯号）的精确度，建立常识库困难太大，但也有人不畏繁琐和艰险尝试着在做。

**613**，利用机器学习（不可解释）算法，又发现十万多个月球陨石坑在嫦娥 5 号挖回珍贵月壤后，中国、意大利、冰岛等国研究人员利用机器学习 AI 算法，联合发布又发现 109000 多个月球陨石坑。该算法用于训练的数据模型来自中国的月球轨道器所提供。

**614**，济南首条全自动无人驾驶地铁通车

近日，济南首条全自动无人驾驶地铁 2 号线通车试运行！采用人脸识别进站、物联网和 AI 帮助乘客选择乘坐车厢、车站智慧大脑启动运行。

**615**，杭州建立民用无人驾驶航空试验区

日前，杭州民用无人驾驶航空试验区建立！构建空地一体未来的交通体系！这是全国第一批城市应用场景试验区，其目标是使无人机在城市低空条件下超视距运行，能够更加多样化、常态化、规模化、全方位构建一套空地一体化的未来智慧交通体系。

无人机在城市空间里的物流、应急保障、医疗配送等场景中的应用。

**616**，自动驾驶将作为未来交通运输领域的重要发展趋势

从交通部获知：2025 年将实现自动驾驶场景化运用，机场、物流等场景或将率先落地。

**617**，上海微技术工业研究院研发出全球目前最快的智能核酸检测仪！

仅需 11.5 分钟就能检出结果，2021 年一季度有望上市投入使用。

这款智能核酸检测仪有三个“芯”：高灵敏图像传感器、温控芯片和 pcr 反应芯片，它们是提升检测速度的关键。

618, 韩国 IITP 研究所近日公布世界 AI 强国排名榜（向全球 AI 科学家问卷调查）：①美国，100 分（作为标准参照），②欧洲，89.5 分，③中国，85.8 分，④日本，⑤韩国。

619, 钟义信：《通用人工智能理论》

## 人工智能范式的革命与通用智能理论的创生

钟义信（北京邮电大学人工智能学院）

**提要：**人工智能的研究取得了不少可喜的进展，也面临着许多紧迫的问题。为了应对这些挑战，学术界提出了各种各样的研究思路。笔者相信，每种思路都有其合理之处，都有可能获得一定的成效。不过，根据笔者的理解，人工智能面临的最深刻最严峻的挑战，是学科和时代的大转变所带来的大阵痛：人工智能范式的张冠李戴。因此，必须对人工智能的范式实施“正冠”：颠覆传统学科范式对人工智能研究的束缚，确立信息学科范式对人工智能研究的规范和引领。实施人工智能范式革命的结果，创生了本文要介绍的《通用智能理论》。

### 一、为什么人工智能的根本出路是范式革命？

#### 1、什么是人工智能？

智能是智慧的子集。人类智慧是人类为了生存发展的目的而不断地运用知识去探索未来定义问题（隐智慧）进而解决问题变革现实（显智慧）的能力。其中的隐智慧严格依赖于人类的目的和思辨能力，只有人类才能拥有。而显智慧则依赖于人类的智能求解操作，因此特称为人类智能。它，可用机器模拟。

根据科学技术的拟人律，人工智能是以人类智能（解决问题变革现实的显智慧）为原型、研究具有智能水平的机器为人类提供智能服务的学科。

#### 2、人工智能研究的现状：局部有精彩，整体很无奈

人工智能的研究，存在三大学派：以模拟人脑结构为导向的人工神经网络学派，以模拟人脑逻辑功能为标志的专家系统学派，以模拟智能系统行为为特色的感知动作系统学派。但互不认可，未能形成合力。

经过数十年的努力，三大学派的研究都取得了一些精彩的成果。如人工神经网络的深度学习，专家系统的机器博弈，感知动作系统的智能机器人等。

但是，另一方面，三大学派的研究更面临着许多问题的困扰。其中最为严峻的挑战包括：它们的理解能力（真正的智能水平）都非常低；它们的通用能力都非常差；至今未能形成人工智能的整体理论。这些问题的严重性在于：

(1) 智能水平低下，就不够资格成为真正的人工智能。

(2) 没有通用的整体人工智能理论，表明人工智能的研究还没上轨道。

系统学的原理表明：整体远远大于部分和。再多再好的部分成果之和，也远远不及整体成果。但是如何才能构建通用的整体理论？至今仍然没有共识。

### 3、人工智能存在问题的总根源：范式张冠李戴

思想指导行动，学科研究活动的指导思想就是学科的科学观和方法论。学科的科学观阐明“学科的本质是什么”，学科的方法论则阐明“学科的研究应当怎么做”。于是，学科的科学观与方法论一起，就决定了学科应当遵循的规范研究方式，称为学科的研究范式，简称范式。

人工智能的研究之所以存在上述这些严重问题，根本原因在于：人工智能是一类开放复杂的信息系统，却遵循了传统物质学科范式的方法论：

(1) 人工智能被分解为结构模拟、功能模拟、行为模拟三大学派，归因于运用了传统学科范式的“分而治之”方法论。对复杂信息系统施行“分而治之”的结果就割断了复杂信息系统各个子系统之间复杂而隐秘的信息联系，而这些复杂隐秘的信息联系正是复杂信息系统的生命线和灵魂。失去了生命线和灵魂的各个子系统，就不再可能合成原来的复杂信息系统！这是现行人工智能研究不能建立“整体理论”的根本原因。

(2) 人工智能系统智能水平低下，归因于运用了传统学科范式的“单纯形式化”方法论。智能的决策能力依赖于对研究对象的形式、内容、价值的全面理解。施行“单纯形式化”丢弃了智能的“内核”：内容和价值因素，而仅仅了解对象的形式，极难做出智能的决策。这是“智能水平低下”的根本原因。

众所周知，“分而治之”和“单纯形式化”是传统物质学科的方法论。它们对传统学科的研究、发展与繁荣做出了伟大的历史性贡献。然而把它们用到作为开放复杂信息系统的人工智能研究领域，就犯了范式“张冠李戴”的大忌！

### 4、人工智能研究范式的张冠李戴，是历史的必然

人工智能范式的张冠李戴并非偶然现象。这是因为，学科的研究活动属于社会存在，它的范式则属于社会意识。由于受到“社会意识滞后于社会存在”法则

的制约，20 世纪中叶信息学科的研究活动迅猛兴起之后，直到如今都没有形成信息学科范式的共识，信息学科的研究便沿用了业已存在而且众所熟知的传统物质学科范式，于是造成了信息学科（含人工智能）范式的张冠李戴。

千年的科学发展，都属物质学科（含材料科学和能量科学），都遵循着同样的物质学科范式，因而没有发生过范式的变革。但是当今却正在发生由物质学科主导的学科体系向信息学科主导的学科体系的历史性大转变：研究活动的大转变在前，研究范式的大转变在后—发生在人工智能由初级阶段进入高级阶段的当下。因此，人工智能范式的大转变（范式大革命）正是这个历史大转变所带来的大阵痛。没有这个大阵痛，便无法完成人工智能研究由初级阶段向高级阶段的转变。这是不以人的意志为转移的客观规律。

## 5、范式革命：人工智能基础理论重大突破的必由之路

既然人工智能存在问题的总根源是“范式的张冠李戴”，那么，解决问题的对症良方就应当是“正冠”：颠覆传统物质学科范式对人工智能研究的束缚，确立信息学科范式对人工智能研究的规范和引领。

那么，什么是信息学科的范式？表 1 示出了信息学科范式的内涵特征。为了便于比较，表 1 还列出了传统学科的范式以及现行人工智能所实行的范式。

表 1 学科范式的比较

事项	科学观	方法论
经典物质学科	机械唯物的物质观 对象：物质客体，排除主观因素 关注：对象的结构与功能 遵守：确定性演化，具有可分性	机械还原的方法论 描述方法：纯粹形式化 判断方法：形式匹配 宏观处置：分而治之
现行人工智能	近准的“物质观” 对象：脑物质，排除主观因素 关注：对象的结构与功能 遵守：可分性	明确的“还原论” 描述方法：纯粹形式化 判断方法：形式匹配 宏观处置：分而治之
现代信息学科	唯物辩证的信息观 对象：主体驾驭的主客互动信息过程 关注：主体目的 遵守：不确定性演化	信息生态方法论 描述方法：形式—内容—价值整体化 判断方法：内容理解 宏观处置：生态演化

通过表 1 三种范式（科学观和方法论）的详细解析和对比，可以十分清晰地看出：作为复杂信息系统的现行人工智能研究，它所遵循的研究范式本来应当是信息学科的研究范式（后者的内涵特征参见表 1 的第三行），但实际上却遵循了传统物质学科的研究范式（试比较表 1 的第 1 和第 2 行）。这就是“人工智能范式张冠李戴”的具体表现。

可见，只有颠覆和摒弃传统学科范式对人工智能研究的制约，确立信息学科范式对人工智能研究的规范和引领，人工智能的发展才能走上正确的轨道。

## 二、范式革命怎样创生通用智能理论？

如上所述，人工智能的范式革命，就是在人工智能的研究领域以现代信息学科的范式置换传统物质学科的范式。那么，信息学科的范式又将怎样规范和引领人工智能的研究呢？

### 1、 总结“学科创生”的普遍规律：范式引领学科创生的全局全程



信息学科范式引领创建“通用智能理论”的全部阶段和历程，可用表 2 所总结的“学科创生规律”来描述。

表 2 学科创生规律

生长阶段	生长模块	模块要素	要素解释
自下而上探索范式的阶段	学科探路	多方摸索	通过长期自下而上多方摸索，总结失败教训和成功经验，提炼学科的研究范式（这一阶段常有“盲人摸象”现象）
自上而下落实范式的阶段	学科范式（宏观定义）	科学观	宏观上明确学科本质“是什么”
		方法论	宏观上明确学科研究“怎么做”
	学科框架（具体定位）	学科模型	基于学科范式的学科全局蓝图
		研究路径	基于学科范式的整体研究途径
	学科规格（精准定格）	学术结构	基于学科范式和框架的学科宽度规格
		数理基础	基于学科范式和框架的学科深度规格
	学科理论（完整理论）	基本概念	基于范式、框架、规格的学科基本知识
		基本原理	基于范式、框架、规格的学科概念联系

表 2 的“学科创生规律”显示，学科的创生需要经历两个阶段：首先是自下而上探索范式的阶段，然后是自上而下落实范式的建构阶段。

探索阶段特点是：各种相关学术背景的研究者们各自进行自下而上的摸索、讨论和争论，总结失败的教训和成功的经验，逐渐提炼出普遍认可且科学合理的学科范式（科学观和方法论）。由于是各自展开的摸索，常有“盲人摸象”的现象，因此，这一阶段经历的时间可能很长。信息学科范式的形成便超越了半个多世纪，至今尚未在学术共同体达成共识。

建构阶段特点是：根据探索阶段所总结的学科范式（即由学科的科学观和方法论阐明的学科定义），就可根据学科定义构筑学科的框架（即学科宏观定义的具体化，包括学科的全局模型和研究路径）、进而拟定学科的规格（即学科规格的精准化，包括学科的学术结构和所需数理基础的水准规格）和最终落实学科的基本理论（即学科理论的完整化，包括学科的基本概念和基本原理）。

从表 2 可以看出，在学科创生的整个过程中，学科的范式自始至终都扮演着最高引领者和规范者的角色：整个探索阶段的任务是为了找到学科的范式，而整个建构阶段的任务则是为了落实学科的范式。

揭示和阐明“学科创生规律”，是构建“通用智能”的第一步。以下将分别阐述笔者团队依照表 2 的规律构建“通用智能理论”的各个步骤和结果。

## 2、研究和提炼“信息学科范式”

从当前的实际情况可以判断，无论是国内还是国际，人工智能学科应当遵循的范式都还处在“探索阶段”，这表现为：各种不同背景的研究人员仍然在按照各自对人工智能学科范式的理解进行着不懈的探索，争论频出，共识鲜有。

由于接受了“整体观和辩证论”思想的熏陶，也由于特殊的学术兴趣、背景和经历，本文笔者早在半个世纪之前就开始密切关注和潜心探究信息学科的科学观和方法论，为形成信息学科的范式做了长期的研究和积累。

1962 年笔者作为北京邮电学院信息论专业的研究生，在研读信息论原著的时候发现它的信息概念只关注了波形，丢了信息的内容和价值这两个核心要素，造成了“信息概念空心化”。于是开始探索、并于 1986 年发表了“形式、内容、价值”三位一体的“信息的综合测度”（即“全信息理论”）。接着，1988 年出版了国内外第一部以“全信息理论”为基础的《信息科学原理》。1987 年又发现人工智能的研究被分解为人工神经网络、专家系统以及后来的感知动作系统三个互不相容的学派。于是笔者认识到，从信息论到人工智能，都受到了传统物质学科方法论“单纯形式化”和“分而治之”的影响，而这些方法论与信息学科（包括信息论和人工智能等）的特点格格不入。

基于这些认识，笔者在 1988 年出版的《信息科学原理》第 10 章就专门探讨了“信息科学的方法论”。此后在每次再版（直到 2013 年第五版）都深化了信息学科方法论的探讨。2014 年，笔者出版了另一部专著《高等人工智能理论》，又在全书的第一篇安排了专门的“高等人工智能的科学观与方法论”探讨。2017 年在《哲学分析》杂志发表了论文“从机械还原方法论到信息生态方法论”。最终，总结提炼成为本文表 1 第三行信息学科范式内涵的标准表述，包括它的科学观和方法论。这是我国关于信息学科范式研究的基本成果。

## 3、范式革命：颠覆旧范式，确立新范式

颠覆传统物质学科范式对人工智能学科研究的统领地位，具体来说就是要在人工智能研究领域摒弃如下的科学观和方法论：

(1) 传统物质学科的科学观：

(1-1) 把研究对象看作是纯客观且遵循确定性演化规律的物质

(1-2) 彻底排除主观因素

(1-3) 研究的目的是阐明物质的结构

(2) 传统物质学科的方法论：

(2-1) 分而治之方法

(2-2) 纯粹形式化方法

确立信息学科范式对人工智能学科研究的规范和引领作用，具体来说就是在人工智能研究领域确立如下的科学观和方法论：

(1) 信息学科的科学观：

(1-1) 把研究对象看作是主体与客体互动的具有不确定性的信息过程

(1-2) 强调主体的驾驭作用

(1-3) 研究目的是实现主体的目标

(2) 信息学科的方法论：

(2-1) 信息生态方法

(2-2) 形式、内容、价值三位一体的整体化方法

可见，信息学科范式与传统物质学科范式两者几乎相反相成。然而，千百年来传统物质学科范式已在人们心中深深扎根，而信息学科范式才刚刚露出嫩芽。因此，要让人工智能研究领域的科技工作者放弃自己所熟悉所深信的传统物质学科范式，绝非易事！不过，非如此就不能取得人工智能的突破与升级。

#### 4、构筑全新“人工智能全局模型”（学科框架A）

信息学科范式的科学观业已指明：人工智能是在主体驾驭下的主体与环境客体相互作用所形成的不确定性信息过程。其中，所谓“主体（特别是人类主体）驾驭”，实际是指：人工智能系统必须接受主体所提出的“求解问题”、主体预设的“求解问题目标”，以及主体所提供的“相关知识”；人工智能系统只能在这个框架下去实现主体所预设的目标。

根据这些思想，可以构筑图 1 所示的人工智能学科研究的全局模型：

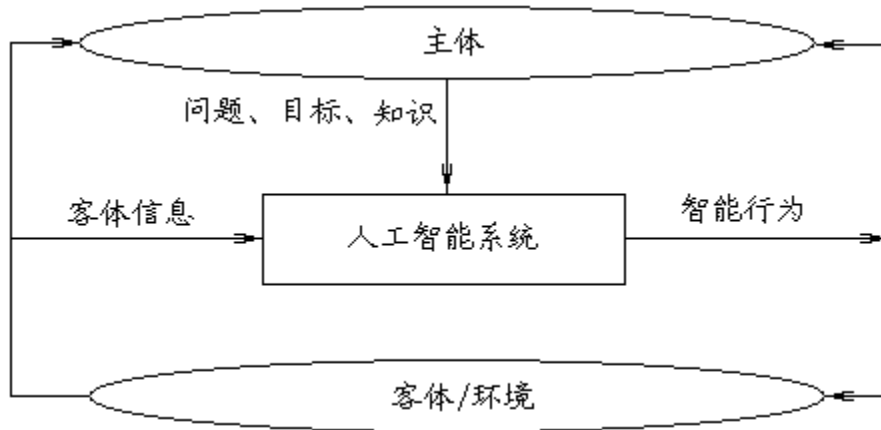


图 1 根据信息学科范式构筑的“人工智能全局模型”

图 1 的模型清楚表明：人工智能的全局模型确实是“主体驾驭下（接受来自主体的问题、目标和知识）的主体与环境客体相互作用（接受环境“客体信息”的作用、针对这种作用产生“智能行为”反作用于环境客体）的具有不确定性的信息过程”。这是真实人工智能系统的正确抽象。

现有的人工智能研究，包括以结构模拟为特征的人工神经网络和以功能模拟为特征的专家系统，都把“孤立的脑”作为全局研究模型的原型。事实上，不接受外部环境客体信息刺激的孤立脑不能产生智能（“印度狼孩”的实验），而不向外部环境输出反作用的孤立脑也不可能检验脑的工作是否有意义。

由图 1 的人工智能模型还可以看出，人工智能系统所实现的，完全是自然主体的目的，而不是人工智能系统“自己的目的”。事实上，人工智能系统由于没有生命，因此不可能有它自身的目的和欲望。它不可能脱离主体的意志自行其事，而只能成为人类主体的聪明助手与善解人意的合作伙伴。

### 5、揭示智能生成机制，开创机制主义研究路径（学科框架 B）

信息学科范式的方法论指明：要按照信息生态演化（既然是生态演化，就不允许把信息剥离成形式化的语法信息，也不允许把信息生态过程割断为封闭的信息过程）方法论来处置问题、要坚持运用形式、内容、价值三位一体的整体化方法来分析问题、要凭借理解来作出判断。

把信息学科范式的方法论与图 1 的全局研究模型相结合，就明确了人工智能系统生成智能的机制应当是在“问题、目标、相关知识”的约束下实现由“客体信息”到“智能行为”的复杂转换，如图 2 所示。

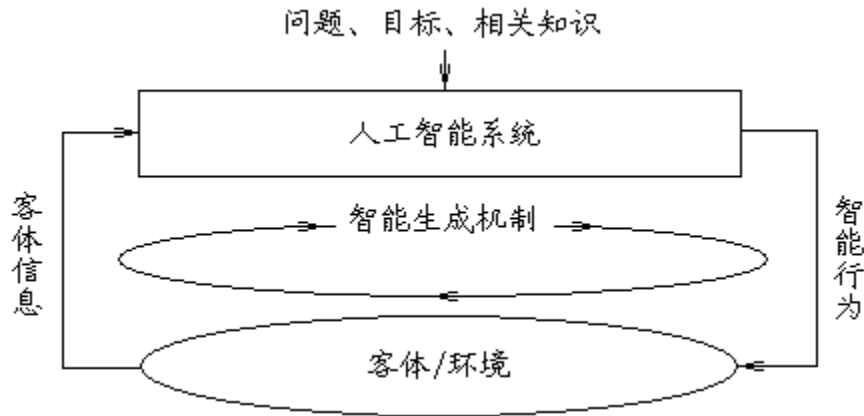


图 2 人工智能系统的智能生成机制

图 2 显示，人工智能系统中的智能生成机制的激励条件，是环境客体所提供并经主体所选择的问题（即“客体信息”），智能生成机制所生成的结果是由客体信息转换而成的“智能行为”，而它所遵守的约束条件则是由主体所提供的预设目标和知识。

智能生成机制的实质是“信息转换与智能创生”，具体转换与创生过程则是：客体信息→感知信息→知识→智能策略→智能行为，如图 3 所示。

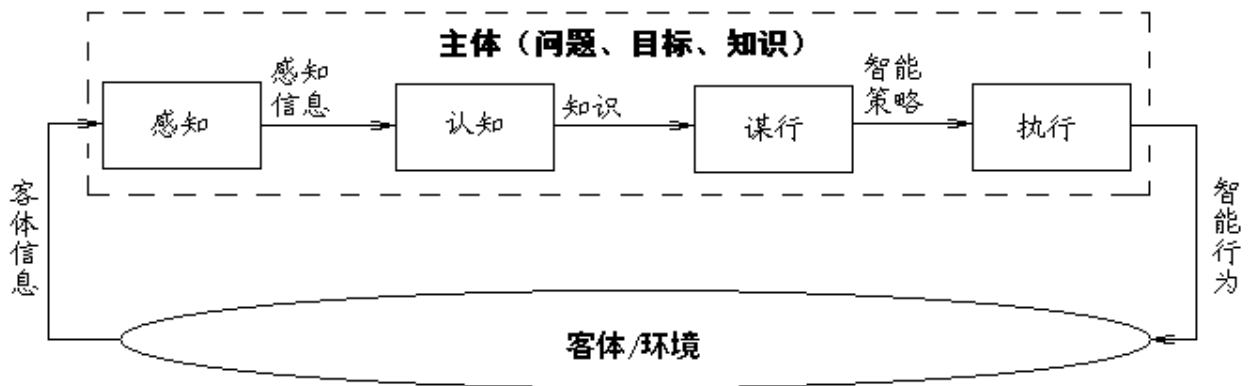


图 3 智能生成机制“信息转换与智能创生”的具体化

可以证明，“信息转换与智能创生”机制是普适性的，不仅适合于各种人工智能也适合于自然智能（包括人类智能），因此，可以名副其实地把它称为普适性智能生成机制，并把它本质称为“信息转换与智能创生定律”。

普适性智能生成机制是人工智能研究的根本路径，以“普适性智能生成机制”

为标志的研究路径称为“机制主义”研究路径。这是与现有人工智能的结构主义路径、功能主义路径、行为主义路径鼎足三分的情况截然不同的全新研究路径。

## 6、重审人工智能的学术结构（学科规格 A）

信息学科范式要求保持学科的学术结构整体性（完整性），恢复人工智能学科的本来面目。遵循完整性的要求，就应把人工智能学科的学术结构理解为以下各个学科群的交叉与综合：

原型学科群：人类学，神经科学，认知科学，人文科学，社会科学，哲学

本体学科群：信息科学，系统科学

基础学科群：生物物理学，逻辑学，数学

技术学科群：微电子，微机械，新材料，新能源

由于传统物质学科范式强调对复杂对象施行“分而治之”，结果就把人工智能学科分解出一些互不相容的分支学科，从而产生对人工智能学科的片面认识和误解。最典型的一种误解就是，仅仅根据专家系统的一家之情，把整个人工智能学科看作是计算机学科的一个分支。这显然不是对人工智能的正确理解。

## 7、重塑人工智能的数理基础（学科规格 B）

遵循信息学科范式的“统一性和整体性要求”，需要改造和重塑人工智能的数理基础，特别是它的逻辑基础和数学基础。

为此，本团队一方面创建了具有可调参数的“柔性逻辑理论”，从而把原来的标准数理逻辑和各种非标准逻辑纳入统一的逻辑连续谱系；同时本团队创建了以因素为基元的“因素空间数学理论”，从而把原来互相独立发展的普通集合论、概率论、模糊集合论、粗糙集合论等相关数学分支和谐地纳入统一完整的人工智能数学理论。

## 8、重构人工智能的基础概念（学科理论 A）

在传统物质学科范式指导下，人们建立了一批人工智能的基础概念。但是，由于接受了“单纯形式化”方法的影响，这些概念只有形式因素而没有内容因素和价值因素，因此基本都是“空心化”的基础概念，比如形式化的“数据”，形式化的“知识”，形式化的“智能”等等。事实上，正是这些空心化的概念，使现行的人工智能系统的智能水平（理解能力）非常低下。

符合信息学科范式理念的基础概念包括：全信息，全知识，全智能等。这里的前置词“全”并非要求“胡子眉毛一把抓”，而是强调“形式、内容、价

值”三位一体的整体化。只有全面了解事物的形式、内容和价值，才能理解事物。

全智能来源于全知识，全知识来源于全信息。因此，智能理论最基本的概念是“全信息”。它的“形式、内容、价值”三位一体整体化体现为“语法信息（形式）、语义信息（内容）、语用信息（价值）”的三位一体。“全信息”的概念具有严格的生成机制和定义方法，如图 4 所示（详细的解释件文献：

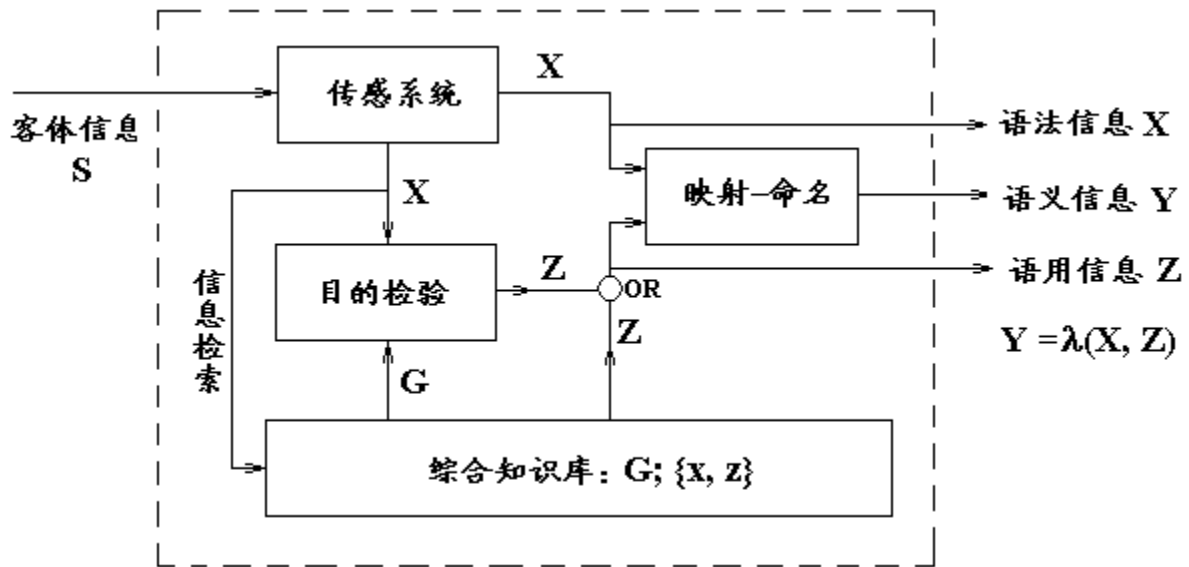


图4 “全信息”的生成机制

图 4 的模型不仅阐明了全信息的生成机制，而且给出了语义信息（内容）的科学定义： $Y = \lambda(X, Z)$ ，其中算子“ $\lambda$ ”代表“映射与命名”的逻辑操作。可见，人们掌握了语义信息，就同时掌握了语法信息和语用信息，就代表了全信息。所以，语义信息是用来“理解”事物的，而不是用来描述事物的统计特性的。

有了全信息的概念与生成机制，全知识与全智能的概念与生成便水到渠成。由此就可以建立“全知识”的知识库（标准的名称是“机制主义的记忆库”），它与传统知识库的根本区别就在于它的知识包含了“形式性知识、内容性知识、价值性知识”，因此可以有力地支持理解。值得指出，这样的“机制主义记忆库”比目前流行的“知识图谱”更为深刻，因而更为优越优越。

总之，只有在“全信息”和“全知识”的基础上，才能具有“理解能力”，才能支持真正的“全智能”。

## 9、深挖人工智能的基本原理（学科理论 B）

信息学科范式强调“信息生态方法论”。因此，最为深刻的人工智能原理就是体现信息生态演化的“信息转换与智能创生定律”。这是一切人工智能和人类

智能系统的本质和灵魂。

正如图 3 所表明的那样，“信息转换与智能创生定律”具体包含：

- (1) “客体信息→感知信息（感知）”的转换原理，
- (2) “感知信息→知识”（认知）的转换原理，
- (3) “感知信息、知识与目标→智能策略”（谋行）的转换原理，
- (4) “智能策略→智能行为”（执行）的转换原理，和
- (5) “误差信息→优化智能行为”（优化）的转换原理。

细心的读者可能已经发现，“信息转换与智能创生定律”的前四项转换原理恰好体现了中华文明的“知行学说”一知（感知、认知）行（谋行、执行）。最后这项“优化原理”，其实是在“利用知行学说来纠正误差”。所以，“信息转换与智能创生定律”完全与中华文明的知行学说交相辉映。

不难证明，“信息转换与智能创生定律 / 知行学说”既是普适性的人工智能生成机理，同时也是人类智能的生成机理。当然，只有应用信息学科的范式才有可能发现这一极其重要的结论。相反，如果恪守传统物质学科的范式，那就只能看到人工智能系统的结构与人类智能系统的结构之间的巨大差别。

值得指出，“信息转换与智能创生定律”的深远意义更在于，它是与物质科学领域的“质量转换与物质不灭定律”和能量科学领域的“能量转换与能量守恒定律”等量齐观的科学定律，它们三者一起就完善了物质、能量、信息三大资源领域的三大科学定律。

进一步的分析可以看到，“质量转换与物质不灭定律”和“能量转换与能量守恒定律”阐明了这两个领域所存在的界限，告诉人们“不能逾越这些界限”。而信息转换与智能创生定律则告诉人们“可以通过信息转换来创生人工智能系统为人类提供智能服务，把人类从体力劳动和有规可循的智力劳动中解放出来，以便更好地发挥人类的创造性能力，实现人类社会的可持续发展”。或者，更简要的说，物质不灭定律和能量守恒定律告诫人们“不能逾越界限”，而智能创生定律则告诉人们“可以创造未来”。

从这个层面上是否也可以说，“信息转换与智能创生定律”对于人类的进步与发展具有更加重大的意义？



## 10、创建通用智能理论（学科理论体系）

综合集成第二节第 1-9 小节的各项成果，特别是其中以“信息转换与智能创生定律”为表征的机制主义研究路径，可以构建既适用于人工智能也适用于人类智能的“通用智能理论”。表 3 列出了通用智能理论的总体特征以及通用智能理论

论总体特征与现有人工智能理论总体特征的对照。

表 3 通用智能理论的总体特征及其与现有人工智能理论的对比

比较事项	现有人工智能理论	通用智能理论
科学观	客观物质，关注物的结构，确定性变化	主体客体互动，关注主体的目的，不确定性变化
方法论	机械还原方法论（单纯形式化；分而治之）	信息生态方法论（形式内容价值三位一体化；整体性演化）
全局模型	孤立的脑模型； 智能系统行为模型	主体驾驭下的主体与客体相互作用 信息过程模型
研究路径	互不相容的研究路径：结构模拟、功能模拟、行为模拟	统一的研究路径：机制模拟（普适性的智能生成机制）
学术结构	计算机学科的一个分支	神经科学、认知科学、信息科学、 社会科学和数理科学等的交叉与综合
数理基础	概率论，数理形式逻辑	柔性逻辑理论，因素空间理论
基础概念	形式化数据、形式化知识、 形式化智能	形式内容价值三位一体的信息、 形式内容价值三位一体的知识和智能
基本原理	未有总结	信息转换与智能创生定律
最终结果	三个局部理论	通用智能理论

由此可以看出范式革命对于智能理论研究的巨大作用和深刻意义。通用智能理论的功能模型如图 5 所示。

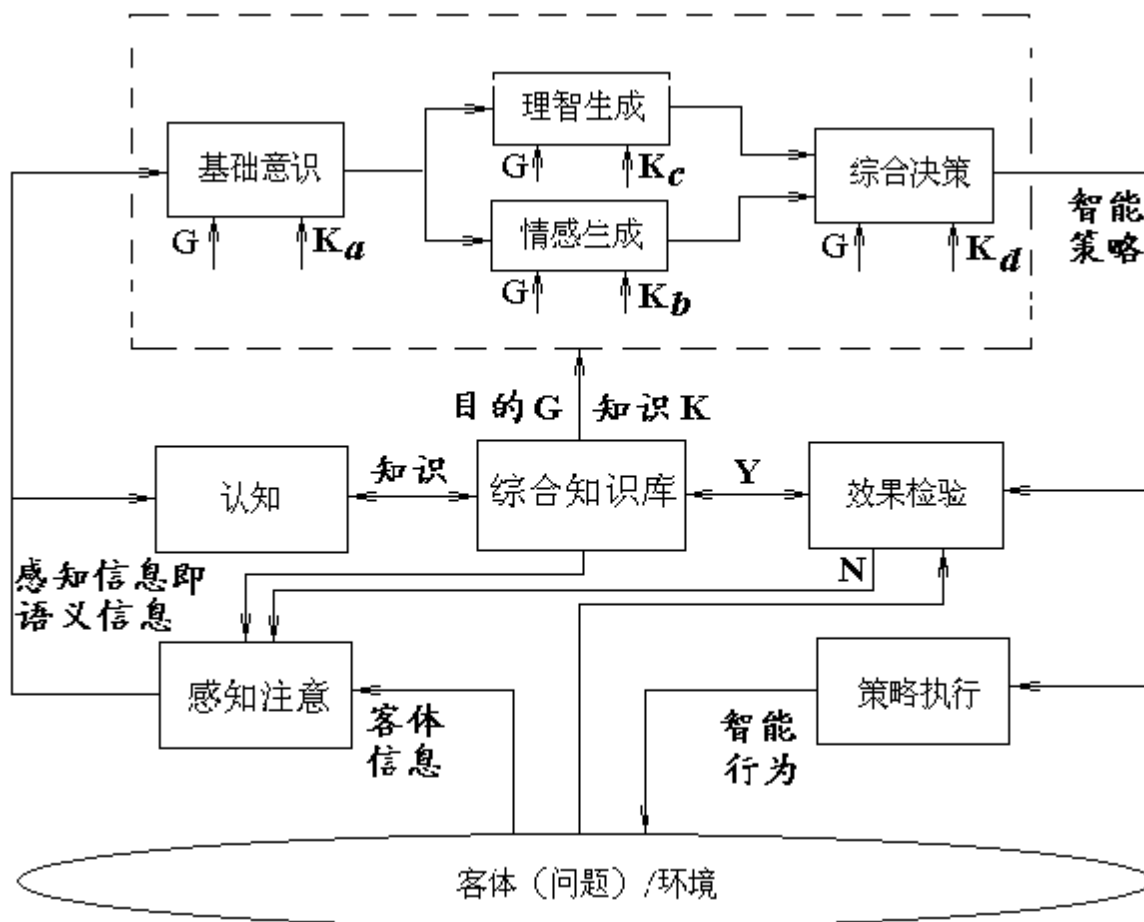


图 5 通用智能理论模型

不难看出，图 5 的通用智能理论模型不但全面体现了信息学科范式（科学观和方法论）的理念，体现了人类学、神经科学、认知科学、信息科学（含智能科学）、系统科学、柔性逻辑理论、因素空间数学理论等科学精神，而且，也展现了人类智能的精髓。特别体现了“物质变精神和精神变物质”的辩证法，以及“人类不断地认识世界和改造世界，并在改造客观世界的过程中改造自己”的人与环境相互作用的理论。

620, 自动驾驶飞行器去年 12 月 28 日从广州起飞，拉开布局城市智慧交通新帷幕

广州开发区交投集团与亿航智能设备（广州）有限公司合作，计划以智能自动驾驶飞行器在交通领域应用为切入点，围绕城市空中交通领域载人级自动驾驶飞行器、智能指挥调度中心、智慧空中物流、空中应急救援等应用，探索激发自动驾驶飞行器应用场景潜力。

## 621, AI 助力艺术

清华大学双聘教授沈向洋发表《AI 助艺术》当题的新年寄语中谈到：

人工智能在创造方面，最近也取得了一些新的进展。尤其是利用深度学习（如生成对抗网络）和语言模型（如 GPT-3）等技术，几乎可以凭空合成逼真的图像和合理的文字。

## 622, 亚马逊云服务（AWS）推出 AI 机器学习服务（即使在不可解释的状态下）

AWS 机器学习服务始自 2017 年，近年来在全球爆发式增长，每年发布近 200 个机器学习服务，覆盖金融、汽车、医疗、工业制造等多元领域。

AWS：机器学习可以说是近代最具颠覆性的技术，2020 年 AWS 已发布 250 个新功能应用，从 Tensorflow 到 PyTorch，从 SageMaker 到 AmazonNeptuneML，从基础架构到顶层应用，并演进到 AISaaS 也会有很多服务。

AWS：计算、算力，针对机器学习仍然在创新，其中的一个重点是选择，AWS 把选择给用户，这些选择永远围绕更合适的性能和成本去推进。

最近 SageMaker 推出一个新功能叫 FeatureStore，一方面可解决特征的一致性问题的，在训练和预测方面都用同一套特征，另一方面能够解决训练和预测中对特征数据存储不同的需求。

AWS：帮助更多的人学习和使用机器学习，将机器学习能力跟数据库 Aurora 嫁接，解决端到端的实际应用问题。

## 623, 学习、评论新一代人工智能

钟、张谈发展新一代人工智能，有共同点：他们都认为在人工智能 60 多年发展中，一直存在着相互竞争的发展途径，但这些代表不同学派的范式途径，缺乏整体观，具有片面性；他们提出的发展第三代人工智能（钟提出发展通用人工智能），应该建立在融合这些不同发展途径的统一范式（统一的模型架构）。也有区别：钟的发展模式是以哲学社会科学和自然科学相结合的理论基础，把结构主义、功能主义和行为主义三大研究途径融合起来，改造成为统一的发展范式；张的发展模式是把符号主义与连接主义两大研究途径融合起来，改造为统一的发展范式，他主张第三代的发

展途径是基于经过改造的统一范式，融合第一代知识驱动、第二代数据驱动，利用知识、数据、算法、算力四要素，研发第三代人工智能（发展安全、可信、可靠与可扩展的人工智能）。钟认为，不是仅靠算法、算力、数据的改善就能解决，也不是把数据驱动与知识驱动相结合就可以了。

其实，关于不同发展途径的争论从 1969 年就开始了（我们在跟贴 150 中介绍了马文-

明斯基和西摩尔-派普特在《Perceptrons》书中的讲述）。神经网络在上世纪 90 年代开始普及，30 多年来，符号主义者与连接主义者争论一直不断，2018 年 12 月，以坚持符号主义的加里-马库斯（Gary Marcus，纽约大学教授）为首的学者，以 Tiwtter 为阵地，向连接主义、深度学习先驱蒙特利尔大学教授 Yoshua Bengio 及脸谱深度神经网络大师 Yann Lecun 重燃了古老的争论。

2019 年 10 月 OpenAI 提出混合模型，2019 年 12 月微软提出神经网络人工智能（neurosymbolic AI），旨在弥合神经表示和符号表示之间差距的模型架构。

回过头来也谈谈钟、张的新见解，要把不同学派融合改造难度不小（从国外符号主义者与连接主义者几十年来争议未解可见一斑），另外，钟、张的新理论至今只是提出一个笼统的概念，尚未见到他们具体的研发方案。还有一点，迄今人工智能带来的繁荣，得益于二大（或三大）学派的发展途径的模式架构，不要轻易将其打碎废除！

## 624，中国已构建了天地一体化广域量子通信网雏形

中科大 1 月 7 日宣布，中国科研团队成功实现了跨越 4600 公里的星地量子密钥分发。

1 月 7 日，中科大潘建伟及陈宇翱、彭承志等与中科院上海技术物理研究所王建宇研究组、济南量子技术研究院、中国有线电视网络公司合作，在《nature》杂志上发表“跨越 4600 公里天地一体化量子通信网络”的论文，证明了广域量子保密通信技术在实际应用中的条件已初步成熟。

《nature》杂志审稿人评论：这是地球上最大、最先进的量子密钥分发网络，是量子通信“巨大的工程成就”。

## 625, 基于跨模态对比学习的多模态理解与生成任务

百度团队提出一个统一多模态预训练架构 UNIMO 能够有效适应单/多模态理解与生成任务

以前的预训练方法只能专注于单模态或多模态任务

——百度 WeiLi, CanGao 等

2020.12.31 发表

以前的预训练方法只能专注于单模态或多模态任务，无法兼顾和相互适应。一种统一模式的预训练架构 UNIMO 可以有效地适应单模式和多模式的理解和生成任务。可以利用大规模的自由文本语料库和图像集合来提高视觉和文本理解的能力，并利用交叉模式对比学习 (CMCL) 将文本和视觉信息对齐到整个图像文本语料库中的统一语义空间中。文本知识和视觉知识可以在统一语义空间中相互增强。UNIMO 显著提高了一些单峰和多峰下游任务的性能。

## 626, 谷歌大脑和卡内基梅隆大学提出 AutoDropout, 能够学习 Dropout 对深度网络进行正则化

——谷歌大脑和卡内基梅隆大学 HiepPham, Quoc V. Le

2021.1.5 发表

Dropout 是谷歌提出的一种正则化技术，用以在神经网络中对抗过拟合。Dropout 有效的原因，是它能够避免在训练数据上产生复杂的相互适应。神经网络经常被过度参数化，可以通过正则化来改善。

使用 Dropout 模式训练目标网络，并将其生产的验证性能用作控制器学习的信号。该方法既适用于 CIFAR-10 和 imageNet 上的图像识别，又适用于 PennTreebank 和 WikiText-2 上的语言建模。所学习的 Dropout 模式具有泛化能力，可以迁移到不同任务和数据集。迄今为止，重用 AutoAugment 和 RandAugment 策略的方法已使 CIFAR-10 和 ImageNet 上的许多最新模型受益匪浅。

## 627, 普林斯顿 (Princeton)、麻省理工 (MIT) 合作: 采用预训练方式实现小样本学习的性能提升

GPT—3 模型通过自然语言提示和一些演示作为上下文，就可以实现令人印象深刻的小样本表现。Princeton 与 MIT 陈丹琦 (DanqiChen) 团队利用小样本的语言模型进行微调，在计算上是有效的。这种方法叫 LM - BFF。方法包括①基于提示的微调，生成提示的新的管道；②完善的策略，用于动态有选择地将演示结合在上下文中。同时还提出了一种系统的方法，分析 NLP 任务的小样本性能。

### 628, Disout 算法 PK Dropout 算法

看到跟帖 626，谈到 Dropout，这是一个谷歌闭源的算法（谷歌已申请专利），谷歌定义 Dropout 为“解决神经网络过拟合的系统和方法”。

这使我想起跟帖 481：“华为开源 Disout 算法 PK 谷歌闭源 Dropout 算法”，重录如下，你为跟帖 626 的补充。

在跟帖 481 中谈到：华为的 Disout 是一种通过研究特征图扰动来增强深度神经网络的泛化能力的方法。Disout 算法超过 Dropout 算法，如在 ImageNet 上训练的 ResNet—50，可达到 78.76%准确率，而 Dropout 训练仅为 76.8%。

### 629, 艾伦人工智能研究所研究员、华盛顿大学计算科学家叶锦才 (Yejinchoi) 创自动知识图谱构建模型 COMET

融合符号推理和深度学习两种 AI 方法攻关常识推理

这在跟帖 457 中已有阐述，现再摘要重录于下：

在强人工智能中，常识推理能力是必备的，这是迄今为止一直困扰人工智能 50 年的难题。

一般解决常识推理有两条路：符号推理和深度学习。但分别学用这两个途径，常识推理无解。

华盛顿大学计算科学家叶锦才团队提出了自动知识图谱构建模型 COMET (CommonsenseTransFormers)，融合了 GOF AI 式的符号推理和深度学习（连接主义）两种人工智能方法，在常识推理攻关方面取得进展，但尚未根本性解决。COMET 在处理超出其内置常识以外（即覆盖性）有不错表现，在解决脆弱性问题上也可圈可点，但深度学习有一个根本性限制：统计不

等于理解（影响了常识推理），要想让计算机建立真正的常识，需要利用语言以外的媒介，如视觉感知或具象感觉（如具人 embodiment），另外还要挖掘神经语言模型的潜力。

### 630, 用于光学神经网络的 11Tops 光子卷积加速器

（11Tops photonic Convolutional accelerator for optical neural networks）

Xingyuan Xu, Bill Corcoran……等研发团队 2021.1.6 发表于《nature》杂志

可查阅 <https://www.nature.com/articles/s41586-020-03363-0>

卷积神经网络（CNN）受源自生物视觉皮层系统的启发，可以提取原始数据的分层特征，降低参数复杂性，提升预测精确度，目前应用于计算视觉、语音识别、棋盘游戏、医疗诊断等，而光学神经网络可以利用广泛可用的光学带宽提升计算速度。

### 631, 中国在 AI、5G 等前沿技术上超过美国？

《华尔街日报》：夸大中国科技实力，危险！ 作者：彼得-考伊等，2021.1.8

美中关系是 21 世纪初的地缘政治大竞争，各方面都将涉及两个大国科技实力。许多美国人忧心中国在人工智能（AI）、5G 通信等前沿技术上已超过我们。此类常见着法显示出对技术一知半解，很易导致误判和政策错误。先从 AI 说起，哈佛大学的格雷厄姆-艾利森称中国是“拥有 21 世纪最有价值商品（数据）的沙特阿拉伯”。这个时髦比喻暗示中国数据多，所以优势大。此类评估犯了两个基本错误，一、机器学习依赖专门的数据集，而非堆积如山的无差别数据点。二、这忽略了收益递减规律，数据输入的无限大供给并不能产生无限好的结果，实际上反而可能降低性能。

除了计算能力，AI 最大驱动力是人才。美国在该领域领先世界，是因为吸引了全球最优秀的科研人员，若美国将这些出色人才的工作当成国家机密封闭起来，或限制他们接收中国学生，就会拖慢他们的脚步，这些人只会把自己的技术带到其他地方去。

华盛顿一些人认为应整垮华为，以恢复美国主导地位。但美国当前政策从根本上误解了决定 5G 竞争力和安全性的因素。5G 发展的驱动力将是半导体、软件系统和云计算——在这些领域世界领军者是美国，而非华为或任何其他中国公司。

美国外交政策制定者不应被华为吓倒，而应认识到中国公司的处境——与 IBM 在大型计算机时代享有的主导地位很相似。IBM 庞大的规模及专有标准和软件曾令竞争对手很难媲美，但随着互联网兴起，它走向衰落。此外，5G 物联网将连接数以万计的设备和软件供应商，大量数据河流跨越国界流动。中国会是个重大安全问题，但只是众多问题中的一个。

权衡是政治家和外交家的工作，他们对基础技术应有深入理解。一个旨在消除对华科技接触所带来的风险的安全战略注定失败，正如我们所看到的那样，甚至许多美国盟友也不会同我们一起打破这种关系。

### **632，融合人工智能不同学派发展路径探索新一代人工智能（通用人工智能）**

在跟帖 150 中，我们看到 40 多年来符号主义者与连接主义者争论一直不断。

2018 年 12 月符号主义者、纽约大学教授 Gary Marcus 又与连接主义者（或深度学习者）、蒙特利尔大学教授 Yoshua 及脸谱首席科学家 Yann Lecun 又燃起了这场古老的争论。

在跟帖 457 中，华盛顿大学计算科学家叶锦才（Ye Jinchoi）提出自动知识图谱构建模型 COMET，企图将符号推理与深度学习融合在一起，攻克常识推理难关，以完善大规模语义网络。

在跟帖 546 中，钟义信教授提出新一代人工智能（通用人工智能）的基础理论，在跟帖 548 中，张钹院士提出迈向新一代人工智能（第三代人工智能）的设想。钟、张均认为符号主义、连接主义等不同学派的发展路径均存在片面性，难以据此单独发展为新一代人工智能，只有将当前人工智能不同学派的发展路径融合改造为统一的发展范式并据此发展，探索新一代人工智能才有希望。

但钟、张团队对如何融合，均未提出具体的解决方案，似乎只有华盛顿大学叶团队为实行常识推理攻关初步提出了具体的融合方案（将符号主义与



深度学习融合），这还是一个初步的融合方案（因为深度学习的统计特征还难以转化为推理或理解）。

### 633，可解释机器学习/深度学习求解程序

#### 一、审题：

首先区分有待求解的可解释人工智能案例是属于整体解释行为，还是局部（个别样本）解释行为？因为不同解释行为将采用不同求解方法。

谈到整体解决行为，属于宏观状态，如全国、行业、地区的规划性或整体性项目，而局部解释行为，属于微观状态，如企业或个体性质的项目。

在审题后，1，按整体或局部解释行为区分，采用不同的求解方法；2，按不同项目内容，开展可解释人工智能的分析、计算程序。

#### 二，分析：

可解释人工智能多用于放贷风险评估项目、或其他决策项目，或反欺诈打假项目，及评定公平、正义或认信项目。

对审议中的不同项目内容，如属发放风险贷款而需进行风险评估的项目；又如属于反欺诈或打假（fake），要求识别假伪欺诈的项目。

分析步骤为：

##### 1) 取特征值或特征函数

如打假项目：

取技术特征识别率、环境干扰影响率、用户自身信用率等。

如放贷项目：

取回款、用户信用、担保等各项特征值。

##### 2) 采用算法

根据各项特征值及单项算法，计算综合特征风险率或综合特征欺诈率。

#### 三，评估：

评估综合放贷风险率，确定风险率可承担红线：在红线以下，表示风险可承担，在红线以上，表示风险不可承担，或也可将风险是不可承担在红线以上分级处之。

评估综合欺诈率，确定风险可承担红线也是如此！

### 634，克服常识推理这块短板

常识推理是语义网络中的短板，克服大规模语义网络的短板，是实现认知智能解决最后一公里问题的关键。

国内外 AI 跟帖 457、589、627、629、634 等都在探索解决这个难题。

处理上下文是扫除妨碍实现常识推理的措施，需要将上下文融合到常识推理的知识图谱中去（可参阅跟帖 589 微软认知服务研究组推荐的措施）；结合语言建模，建设知识库（常识库）、关键路径、图关系网络等，建设 DEKCOR 模型（参阅微软）、建设自动知识图谱构建模型 COME T（参阅华盛顿大学）。

**635**，赛昉科技在上海发布（2021.1.13）：全球首款采用 Linux 操作系统的 RISC—V AI 单板计算机（星光单板机——Beagle—V）

该产品采用惊鸿 7100 系列 AI 视觉处理芯片 SoC，搭载 4 核 CPU，芯片面积 85mmX70mm，工作频率 1.5GHZ，内存 4G、8G，共享 2MB 的二级缓存，支持 Linux 操作系统

该产品可用于工业机械、数据中心、人工智能（AI）、边缘端云计算等领域。

星光单板机的出现将推动全球 RISC—V 生态的发展，推动 RISC—V 技术和产品加速迭代。

该产品是由赛昉科技（StarFive）、矽递科技（Seed）和 BeagleBoard 三方共同开发的。

**636**，评钟文

看到跟帖 546、547、559、559 附件 569 发表钟义信教授关于构建下一代强人工智能（钟称：通用人工智能）的文章，概括来说钟文认为：

1)，在人工智能发展历史上有三大学派，即：①基于数据统计连接系统的结构主义，②基于逻辑推理符号系统的功能主义，③基于模拟智能生物行为感知动作系统的行为主义。在人工智能发展中，各自取得不少精彩成果，但三派争霸无法统一，其发展路径均存在片面性，无法通往下一代强人工智能领域。

2)，通往下一代强人工智能（即欲进入通用人工智能领域），需从突破性变革的学科范式出发（人工智能范式的颠覆性变革催生了通用人工智能的基础理论）。

3)，钟教授提出了构建通用人工智能学科范式的基础理论，由变革并融合三大学派的不同发展路径演变而成。

现在的问题是：如何对三大学派的发展路径进行颠覆性变革和相互融合，以实现统一的学科范式，支持构建下一代强人工智能（通用人工智能）？似乎尚缺具体的解决方案！

### 637，评张文

看到跟帖 548 发表张钹院士等关于构建下一代强人工智能（张称：第三代人工智能）的文章，概括来说张文认为：

第三代人工智能的理念及其发展途径是融合第一代人工智能（符号主义）的知识驱动和第二代人工智能（连接主义）的数据驱动的人工智能范式。把知识驱动和数据驱动结合起来，同时利用知识、数据、算法和算力等四要素，建立可解释和鲁棒的人工智能理论与方法，构建更强大的人工智能，即发展安全、可信、可靠和可扩展的人工智能技术。

张院士提出构建第三代人工智能的范式，如何融合（不同发展路径的融合）？如何结合（两驱动）？如何利用（四要素）？如何构建大规模语义网络（知识图谱），赋予其足够能力可支持认知智能得以实现？似乎尚缺乏具体的解决方案！不能一句话就说过去！

### 638，谷歌开发语言可解释性工具 LIT

——Jeff Dean

为了更好地理解语言模型的行为，谷歌 2020 年开发了语言可解释性工具（LIT），这是一个可以更好地解释语言模型的工具包，使得交互式探索和分析语言模型的决策成为可能。

### 639，清华—工程院联合发布《人工智能发展报告 2020》

清华大学智能研究院与清华-中国工程院知识智能联合研究中心联合发布中国《人工智能发展报告 2020》，其核心发现主要有以下十点：

- 1, 总结了过去十年的十一人人工智能研究热点,
- 2, 过去十年有五位人工智能领域学者获图灵奖,
- 3, 顶刊获得最多奖项是计算理论、安全与隐私和机器学习,
- 4, 人工智能最高引文的头部主题为机器学习、计算机视觉领域,
- 5, 中国在多个相关领域紧跟美国处于世界前列, 在多媒体和物联网领域超过美国, 部分领域还需追赶,
- 6, 全球范围内中国人工智能学者数量占 9.8%, 美国 62.2%,
- 7, 清华大学是唯一入选全球人工智能高层次学者数量 Top10 的中国机构,
- 8, 京津冀、长三角、珠三角地区为国内主要高层次人工智能人才聚集地,
- 9, 过去十年中国专利申请量世界第一, 是第二名美国的 8.2 倍,
- 10, 人工智能未来重点发展技术方向包括: 强化学习、神经形态硬件、知识图谱、智能机器人、可解释性人工智能、数学伦理、知识指导自然语言处理等。

#### 640, 评《人工智能发展报告 2020》

看到由清华与工程院专家撰写的中国《人工智能发展报告 2020》, 长见识! 但这个报告似偏于论文, 不够全面, 现归纳我们国内外跟帖所谈, 作一些补充、修改, 供参考!

1, 国内人工智能通常忽略处于不可解释状态的机器学习/深度学习, 其实正是它支持了今天人工智能的繁荣。上世纪末抗生素产生了抗药性, 麻省理工的科学家通过深度学习模型, 对 2300 多种抗生素药物进行训练, 研发出能解除耐药性、杀死耐药的超级细菌的全新抗生素, 谷歌的科学家利用深度学习训练蛋白质, 研发用于基因医疗的 AlphaFoldv1、2, 最近又研制出抗疫药物, 又如, 英国基于机器学习研发了全球首款由人工智能控制的吸气式高超音速六代机——“暴风雨”战斗机, 再如, 国内外一批科学家基于机器学习筛选用于研发新材料、新药物等等。

2, 自动驾驶、无人驾驶发展很快, 国内外在 L4 路况下研发自动驾驶、无人驾驶竞争激烈, 主要是两个方案: 一是单车智能方案, 通过在车上安装多个雷达传感器、实现对路况环境感知, 提出颠覆性自动驾驶行为预测算法模型, 二是车路协同方案, 在路也配置、安装激光雷达传感器, 由车辆与路边交互而完成对周围环境的感知(这个方案人称“中国方案”), 自

自动驾驶、无人驾驶现已开始形成相当规模的行业，在美国，围绕 Waymo、Zoox、Cruise、ArgoAI、Aurora 硅谷五大家族由几十家企业组成了行业，在中日专，已不是百度一家独大！有腾讯、蔚来、四维图新、小马智行、北京三快、领骏科技、华为、滴滴出行及若干传统汽车厂，其中一些企业也积极参加国际竞赛，而且单项指标或综合指标创优，国内行业现在也有十几家。在目前形势下，自动驾驶、无人驾驶企业如不在创新上下大力、不在建设生态上下大力，优胜劣汰将很现实！

3，谈到人工智能未来重点发展技术方向，报告中提到的可解释性人工智能、神经形态硬件、知识图谱也是我们人工智能国内外跟帖所推荐的，这里要强调挖掘不可解释的机器学习（及深度学习强化学习）潜力，把可解释人工智能改为可解释机器学习，并列为未来重点发展的首位，将知识图谱改为大规模语义网络，目前依托它来实现认知智能（或可解释的人工智能）还差最后一公里，需要在知识指导下提高语义的理解能力、解决常识等短板问题。将神经形态硬件改为基于异步脉冲神经网络（神经拟态硬件）的神经拟态计算系统。有人提出，作为人工智能未来的发展方向，是要探索创建新一代（或第三代）人工智能（或通用人工智能）的问题，如此可能要很长时间，是否列上可权衡考虑。

#### 641，可解释人工智能：工程观点

FatimaHussain 等 SMIEEE, EkramHossainFIEEE

2021. 1. 10

深度学习的显著进步激发了人们在几乎每个领域使用人工智能技术的热情。但是，这些算法的不透明性，使其在安全关键型系统中的应用产生质疑。

“可解释性”维度不仅对于解释黑盒子算法的内部工作至关重要，而且还增加了问责制和透明性维度，这对于监管机构、消费者和服务提供商至关重要。可解释人工智能（XAI）是将所谓的黑盒子算法转换为白盒子算法的一组技术和方法，这些算法获得的结果以及该算法采取的变量、参数和步骤达到所获得的结果，是透明且可解释的。

以自动驾驶汽车为例，讨论 XAI 在其不同组件（例如对象检测、感知、控制、动作决策等）中的应用。这项工作是一项探索性研究，旨在确定 XAI 领域新研究途径。尽管 XAI 在基线 AI 模型中添加了必要的功能，但评估

与解释相关的其他功能的 AI 模型的性能也非常重要。必须考虑新功能可能会产生开销并可能影响准确性。

#### 642, 可解释人工智能和人类互动

美国陆军研究实验室和南加州大学创意技术研究所合作发布

如果人类无法理解同上下文进行有效交互, 人工智能的优势无法体现, 基于可解释人工智能 (XAI) 允许人工智能解释决策就可以缓解这种问题。

本文利用一些需求案例, 探索对运营计划人员 (高级分析师) 被虚拟分析师代替, 它们通过分析、搜索和呈现目标系统来实现需求。

通过将 XAI 集成到现实世界当中的双向 workflows, 能够创建计划并简要说明指令, 一旦出现问题, 就可以进行分析、实现理解。通过分析员共同努力, 对威胁、脆弱性、事件等以应对未来可能出现的攻击。

#### 643, 哈佛大学发布《机器学习药物研发》教程

2021. 1. 7 发布

药物发现和开发是一个漫长和昂贵的过程。它通常始于分子和靶标的实验发现 (即从头设计药物), 然后在进行临床测试之前, 先对细胞系、类器官和动物进行体外实验来验证发现。从发现新药到获法规批准的整个过程可能需要长达 12 年的时间, 成本高达 28 亿美元。此外, 每个药物开发阶段都存在巨大的不确定性 (成功率 1: 5000)。机器学习方法已经为解决这些挑战并加速药物开发的有前途的工具。

本教程涵盖药物有关关键任务: ①合成预测和新药设计, ②分子性质预测, ③虚拟药物筛选和药物靶标相互作用, ④临床试验招募, ⑤药物再利用, ⑥药物不良反应和多元药物食物反应。

#### 644, 可解释性机器学习可信教程

上交大张拳石在 IJCAI2020 会上介绍可解释性机器学习可信教程 (2021. 1. 11)

深度神经网络 (DNN) 无疑为计算机视觉、计算机语言学和人工智能的广泛应用带来了巨大的成功。但是, DNN 成功的可信赖性以及 DNN 对对抗攻击的适应力仍然很大。在可解释的人工智能范围内, 对引人注目的但有争

议的话题。相关问题包括①网络特征可信度的量化，②DNN 解释的客观性、鲁棒性，语义严格性，以及③可解释神经网络的可解释性的语义严格性等。教程旨在将关注人工智能的可解释性、安全性和可靠性的研究人员、工程师以及工业从业人员。

本教程对当前可解释 AI 算法的优势和局限性的批判性讨论提供了新的前瞻性研究方向。预计本教程将对关键的工业应用产生深远的影响，例如医学诊断、金融和自动驾驶。

**645**，如何看待人工智能的发展方向？谷歌作为全球人工智能发展重镇之一，其看法和做法值得我们借鉴。现将谷歌高级副总裁 JeffDean 代表 Google Research 最近发文的总结（谷歌人工智能 2020 年的发展成就，同时也展望 2021 年接下来的工作重点）简述如下。

他列出 2020 年谷歌人工智能在 10 大领域的发展成就：

#### 1) 新冠病毒和健康

基于机器学习算法，研究对新冠病毒的调研、检测、预防、诊治，以及研究新冠病毒流行对健康与经济的影响。

#### 2) 天气、环境与气候变化

基于机器学习算法，研究天气、环境与气候变化。

#### 3) 可访问性 (Accessibility)

例如采用机器学习方法，帮助视力受损用户识别包装食品。

#### 4) 机器学习在其他领域的应用

此处以脑机接口举例，研究机器学习与神经科学联系，研究果蝇的脑组织如何运作。

#### 5) 负责任的人工智能

可解释的机器学习，谷歌开发语言可解释性工具 (LIT)，这是更好地解释语言模型的工具包，使交互式探索和分析语言模型的决策成为可能。

#### 6) 自然语言理解

#### 7) 机器学习算法 (深入研究)

#### 8) 强化学习

#### 9) AutoML

这是一个非常活跃的研究领域，将系统地重塑机器学习算法

#### 10) 更好地理解机器学习和模型

随着神经网络被做得更宽更深，训练得更好，泛化得更好的背景下

在上述 10 项工作中，机器学习是其重点，大多数项目都是基于机器学习的，对机器学习的算法与模型要加深理解、改进、创新。

2021 年谷歌人工智能的工作目标：一是以关键应用场景为导向，二是考虑 2020 年工作的延续，三是要对机器学习加深理解积极运用，四是根据人工智能国际形势发展有所创新。