**女娲智脑项目技术扩展介绍**

1. **与现有技术的横向对比**
2. **为什么深度学习在“理解领域”不能work**

深度学习虽然已经在视觉识别、无人驾驶汽车、语音识别和其他众多方面产生了革命性的影响。但是，深度学习技术仍面临巨大的问题和挑战：

（1）易受欺骗的问题：将一个小小的数字化日光贴纸贴在图像的某个角落，就可能导致一个经过了数万次训练人工神经系统输出错误的结果，这种被谷歌大脑（Google Brian）团队称之为“对抗攻击”的现象，是普遍存在的。

（2）严重的低效率问题。Hinton（Geoffrey Hinton深度学习创始人）说“让一个孩子学会认识一头母牛，并不需要说一万次牛（10000-cow问题）”，而这却是深度学习系统中经常需要的数字。人类通常仅从一两个例子中就可以学会新概念。

（3）非标记、不透明问题。一旦训练好了一个深度学习系统，就无法确定它是如何做出决定的。“在许多情况下，即使AI得到了正确的答案，但也是不可接受的。”（David Cox，麻省理工学院与IBM联合实验室的计算神经首席科学家）。例如，一家银行使用人工智能来评估你的信誉，然后拒绝给你一笔贷款：“在许多州，有法律规定你必须解释原因，但使用深度学习，你解释不清”。无人驾驶汽车，它们使用类似的深度学习技术进行导航，但也造成了广为人知的灾难和死亡事故，而原因却往往“无法解释”。

（4）AI缺乏常识。深度学习系统是非标记的，可能是识别像素间的某种模式，但是他们无法理解模式的含义，更不用说理解它们产生的原因。DeepMind的AI研究员Greg Wayne：“当前的系统根本不能知道沙发和椅子是用于坐着的。”

（5）迁移困难。一个下围棋的网络，无法下象棋；一个识别人脸的网络，无法识别一张桌子

（6）深度学习在感知层面已经有了显著进步，但在理解人类的认知层面还是有重大问题。

\*\*以上引自《深度学习的黑暗时代已来临？》

综上，深度学习等技术在理解领域已被证明“走不通”！

深度学习不是“语义的”——基于已有答案库的抽取，不知道含义；

深度学习不是“可理解的”——基于黑盒概率连接值，人类读不懂；

深度学习不是“无监督、自动的”——基于人类标注（指导），无法自动化。

深度学习反映的是概率统计关联性，找出来的并不是本质上的特征、语义上的特征；找出来的关系，也并不都是因果等逻辑关系，而是关联关系。

也就是说深度学习区分物体的依据是大数据的重复模式，而人类大脑区分物体的依据是语义上的特征，是“小数据”的，两者有本质的区别。

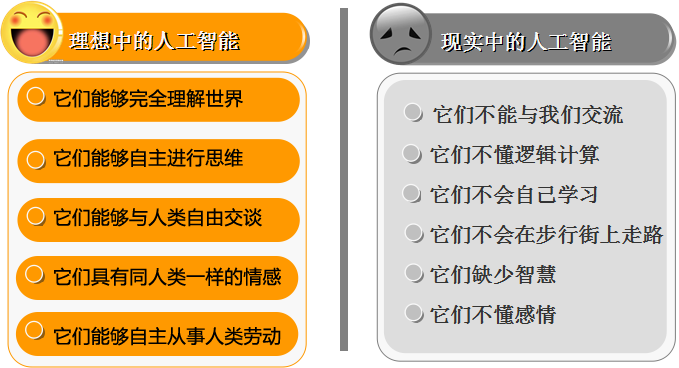
\*\*以上根据中国科学院院士，教授，博士生导师，智能技术与系统国家重点实验室名誉主任 张钹 文章整理

1. **为什么基于乔姆斯基的语法规则的也不行？**
2. 小孩子不懂语法，但可以思维！
3. 语法规则挂一漏万，千差万别，复杂度指数级别！
4. 语法不能产生语义，例如：小明跳了5000米远，符合语法，但语义不对！
5. **为什么知识图谱目前还不能实现强AI？**

知识图谱是最接近人类思维和表达方式的一种人工智能，但是：

1. 知识图谱的关系是固化的，不是“天然”的
2. 知识图谱没有终点，有输入与系统知识比较无效，无输入状态下更是无法取舍；
3. 知识图谱没有深度：牛有腿-牛组件腿-牛组件牛腿
4. 知识图谱没有内外结构
5. **“人工智能”更像是“人工智障”的根本原因——不“理解”**
6. 不理解“对象”，就无法对人类世界进行建模，进而进行分析；
7. 不理解“关系”，就无法对事物之间进行关联，进而产生新的知识；
8. 不理解“动作”，就无法让静态的模型动起来，进而产生“语义”；
9. 不理解“上下文”，就无法对多轮对话进行串联，进而理解用户真正的意图；
10. 不理解“观察者”的内外结构和桥梁作用；
11. 不理解“我”，就无法产生真正的情感，进而与用户进行无障碍交流；
12. 不理解内部函数，就无法形成信息处理的规则及流程，并可根据反馈结果进行优化；
13. **技术领先度**

从2006年开始，项目创始人经过十多年的探索，创造性地提出“基于认知关系的可拓形式化表征知识体系”、“跨领域现象逻辑表示及计算分析模型”、“正态分布式最大似然层次矩阵的聚类与降维计算模型”、“基于二相（多相）纠缠状态的对象转化表示及计算模型”、“基于生命特征的情感表示及情感计算系统”、“基于评判的自适应共振阈值网络”、“自激励式无监督问题求解计算模型”、“基于最大熵及元知识与认知知识关系转化”、“超临界状态的离散问题假设检验”等多项理论，在世界模型、语义表达、动作流程、元知识模型、逻辑生成与运算、自动学习、情感计算等方面取得了多项实质性突破，也对传统的语法学、NLP、图论、人工神经网络（深度学习）、对象表示、本体论、哲学、逻辑学等领域进行了颠覆性的创新。实现了在语义层面、语境层面对自然语言完全理解，包括集合、时空、数字概念，特殊语境指代词、同义词以及上下句语义等。



**AI的愿景与现实对比图**

1. **突破了传统世界模型表示的局限性**

女娲智脑全面突破现有技术：基于“小数据模型”，实现了“基于世界的理解，基于理解的世界”。传统知识表示分为符号表示法和网络表示法。符号表示法包括一阶谓词、产生式规则、框架表示法、有向图等；网络表示法包括语义网络、神经网络（深度学习）。而女娲智脑创造性地研发出了特有的模型，兼顾了这两类知识表示的优点。可以灵活、完整地对世界建模，并且实现了可拓展的动作网络及深度语义网络，具有动态性、可扩展性、可自生成性，是目前最佳的世界理解（UoW）模型。

1. **基于逻辑网络的归纳、泛化、推理、类比技术。**

使用了先进的领域管理技术，可以使知识的领域迁移和推广轻而易举。同时，使用先进的技术对对象进行赋义，可进行线性与空间向量的自由转换。基于上下文、特征向量及个性化的关系抽取技术，这些都使传统技术上很难实现的归纳、泛化、推理、类比在系统中轻松实现。

1. **解决了传统的知识图谱/网络会导致“全脑激活”的问题**

传统的知识图谱/网络（有向图）只有信息的起点，没有信息的终点，例如：信息可以由“苹果”关联到“红色”，再关联“气球”，再关联到“会飞”，再关联到“鸟”等……没有信息终点，最终导致将知识库中的所有概念激活，也就是所谓的全脑激活。而使用激活度递减方法，也由于知识体系的不同，而导致不确定性增加，从而导致系统结果错误！女娲UoW知识表示模型采用特定的结构，突破了几百年没有变革的图论，避免了信息全脑激活。

1. **可以表示丰富多彩的世界关系**

传统知识表示法中，不管是用符号还是用对象，其表示的关系是写死的，例如，传统一阶谓词、语义网络方法中的关系、“is-a”或者“is-part-of”这种关系的用法和内涵都是固定的。而在女娲核心引擎中，所有关系都是随着知识的增加，通过理解、生成、学习得来的。

1. **突破性实现了真正的情感计算**

女娲智脑创造性的提出了“情感计算模型”，根据外界输入信息，情感系统提取记忆信息中的情感因素，按照“女娲智脑”情感模型进行建模，区分对象、事件和动作并进行处理。计算结果会进入情感空间进行处理，形成当前时刻的情感状态。

情感会影响知识库中的计算，影响对记忆的分析，影响归类、反思、联想等引擎的计算结果，进而影响行为系统表现出来的处理。

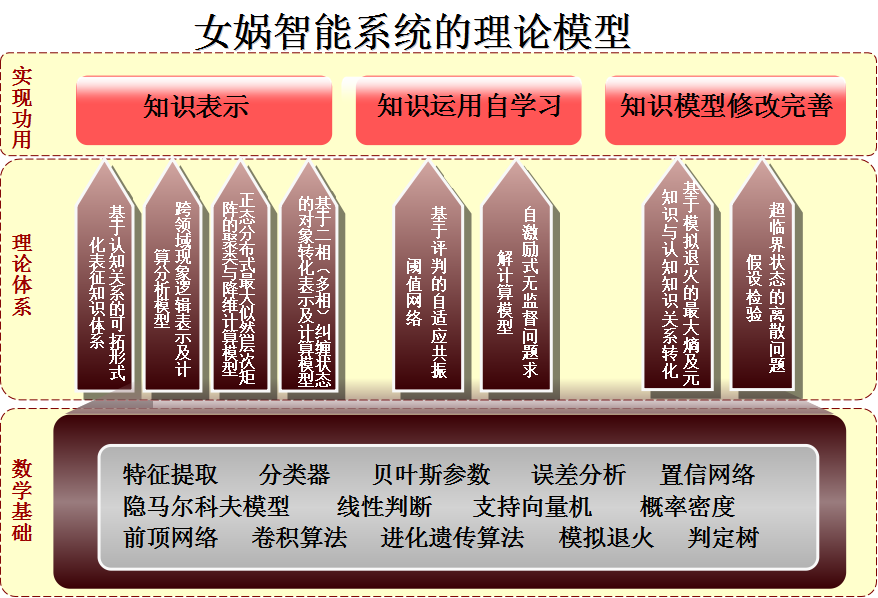
情感有累积效应，会随着时间的推移影响系统的性格，甚至幽默等属性。

1. **可以实现“我”的表示**

传统人工智能的一大难题就是无法表示“自我”，只能就其选择的领域以旁观者或者相对智能信息库的形式发挥其价值，不能真正区分“我”与其他对象。女娲智能引擎可以表示“我”，有自己独特的记忆和知识，有自己的思考、逻辑甚至主张。

1. **可以实现自主学习**

传统人工智能不能实现自主学习，只能实现简单领域知识的获取和按限定规则的整理和“学习”，但并不是真正的学习。女娲核心引擎可以自我学习，学习内容不受门类限制，所有外界信息都是学习的来源；学习的结果可以影响自我对外界的判断和认识，已有知识的改造没有局限（类似人类对外界的认知，知识可以全面改造、发展和更新）。通过建立可操作、使用知识的元知识、元方法库，女娲系统能够像人类一样完成形成知识、存储知识、调用知识、使用知识进行搜索和推理、归纳和演绎知识等任务。



1. **实现知识与“神经网络”的完美融合**

女娲系统突破了知识表示与神经网络的界限，创造性地提出了“模糊的归模糊，清楚的归清楚”的技术指导方针，在感知、统计领域，使用深度学习这样的模糊算法；在逻辑与认知领域，使用女娲UoW知识表示系统。同时，我们将认知计算的结果送入到深度学习网络中，指导深度学习网络的模式建立，通过提高训练速度和效率来缓解深度学习的10000-cow问题。并且它可以使网络更不容易受到对抗性攻击，因为代表一个物体的是系统，而不仅仅是像素，它不会被一点噪音或一个无关的贴纸轻易地干扰。两者较好地融为一个网络，并实现相互转换，使系统主动性有了质的飞跃！