评人工智能如何走向新阶段？
（兼谈国内外跟帖评论）

陆首群 2020.12.17

国内外AI跟帖留言
（552条~584条跟帖）

第七集

中国开源软件推进联盟
China OSS Promotion Union

人工智能国内外跟帖评论大纲

⊙评人工智能向何处去？

⊙机器学习/深度学习/增强学习

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

⊙基于异步脉冲神经网络的神经拟态计算系统

⊙依托大规模语义网络（知识图谱）支持，破解认知智能解决方案

（还差最后一公里）

⊙脑机接口的理论和实践

⊙下一代通用人工智能

（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质疑的质疑

评人工智能向何处去？

（兼谈国内外跟帖评论）

陆首群 2020.12.17

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 2020 年 12 月 17 日（近一年半时间），我们公布了来自国内外的 584 条跟贴，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，）真有点出乎意料！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。

在已发表的 584 条跟帖中，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从知识工程出发，依托大规模语义网络（知识图谱）的支持，破解认知智能解决方案，④脑机接口的理论和实践。目前①已成为全球（特别是美欧）的热点，并已有所突破， ② 已见亮点（国内外已有先例）， ③ 还差最后一公里（世界性问题），

④ 国内外已有几十例试点（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，有关专家怀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，分别提出发展下一代人工智能的解决方案。他们将他们的解决方案汇集到跟帖中来讨论。他们提出的方案尚属于构想，离实效还差得很远，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，其结论将拭目以待！

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，时代之声呼吁拥抱开源，我们需要研究开源与新时代的关系。过去我们虽然论证开源是人工智能、基因科学与深度信息技术的基础，量子计算代表算力，今天和今后的量子计算应该拥有像人工智能及人脑一样的自学和思考能力，量子计算也离不开开源。

在今后发布的跟帖中，我们也将收集反映“新时代”的风貌。我们研究了针对“人工智能向何处去？”的跟帖评论要不要继续办下去？回答是：既然受到国内外欢迎，就应办下去！本集是人工智能国内外跟帖评论的第七集（552 条-584 条），我们还将办好“人工智能国内外跟帖评论续集”，从第八集（或从 585 条跟帖）开始！

国内外 AI 跟帖留言（552 条-584 条跟帖）

552. 无人驾驶与自动驾驶

无人驾驶与自动驾驶技术属于人工智能技术范畴，近年来有长足进步。无人驾驶与自动驾驶的实现与路况密切有关。

根据国际自动机工程师学会2004年制定的“无人驾驶与自动驾驶技术界定标准”，将路况分成 5 级:

0	1	2	3	4	5
No Automation 人工驾驶	Driver Assistance 辅助驾驶	Partial Automation 半自动驾驶	Conditional Automation 高度自动驾驶	High Automation 超高速自动驾驶（人工接手）	Full Automation 全自动驾驶（完全无人）

根据美国高速交通安全管理局（NHTSA）规定，无人驾驶技术也是分 5 个等级（见跟帖 518,538):

L0 级，是人工驾驶

L1 级，需要人类驾驶员（辅助驾驶）

L2-3 级，可实现半自动无人驾驶

L4-5 级，可实现全自动驾驶

L4-5 级区别在于: L4 级需要在特定的道路和天气下的路况，

L5 级可以适应全路况。

目前全球无人驾驶与自动驾驶技术正在主攻 L4 等级路况下行驶无人驾驶与自动驾驶。

在全球还未能实现首款 L5 等级（或全路况）的自动驾驶。

553. 欲在 L4 等级路况下行驶，谷歌等公司 Alphabet 旗下的 waymo 公司提出了颠覆性的自动驾驶行为预测算法模型，即提出一个抽象化认识周围环境信息的 Vector Net,用向量来表达地图信息和移动物体，并在该模型向量间加入了语义关系以帮助进行行为预测，使 Vector Net 具备很强实用性，提升自动驾驶行为程度，训练自动驾驶汽车在现有城市环境中 L4-L5 等级路况条件下进行导航（见跟帖 327、328、486、495）。

百度尝试提出另一种思路，在 2019 年美国的一次自动驾驶技术的讨论中，提出了自动驾驶“中国方案”，即在对现有城市数字化改造的同时根据导航提示传递更丰富的道路信息，在路侧建设完善的激光雷达和传感器系统，根据导航提示传递更丰富的道路信息，以支持实现在 L4 等级路况下的自动驾驶行为（可查看跟帖 145）。今年（2020）在长沙举办首届 Apollo 生态大会上，探索城市、自动驾驶、技术和车厂三位一体（由长沙市提供公共资源，由百度提供自动驾驶技术和生态，由一汽红旗负责整车设计），期望走出一条中国特色的自动驾驶之路（可查看跟帖 118）。

探索 L5 等级路况下的自动驾驶:

在 2019 年末在美国的一次讨论中，全球自动驾驶专家们认为，从技术上看，城市中道路的实时驾驶环境（L5 等级全路况）比路侧导航更复杂，可能出现极端危险情况也明显增多，探索 L5 等级路况下自动驾驶行为难度将更大。

554. 类脑芯片（或神经拟态芯片）几例:

①IBM True North 芯片，2014 研发、2017 发布 。

28nm 制程工艺，芯片尺寸 100 ㎟ ，每颗芯片 4096 个内核，每个内核支持 100 万个神经元、2.5 亿个突触每颗芯片支持 41 亿个神经元、1 万亿个突触。

芯片集成： 48 颗芯片，由芯片集成组成的神经网络: 2000 亿个神经元组网相当于普通老鼠大脑。

②英特尔 Loihi 芯片，2017 发布。

14nm 制程工艺，芯片尺寸 60 ㎟ 。

每颗芯片支持 125 个内核，每个内核支持 1000 个神经元、25 万个突触。

每颗芯片支持 12.5 万个神经元、1000 万个突触。

由芯片集成组成的神经元网络: 1 亿个神经元组网，相当于小型哺乳动物大脑

Loihi 芯片训练后能效提高 1000 倍。

③清华大学 天机芯 （TianJic），TianJic-1 2017 发布。

TianJic-2 2019.7.3 在 Nature 上发表

TianJic-3 尚在研制中

28nm 制程工艺，芯片尺寸 14.4 ㎟ 。

每颗芯片 156 个内核，每个内核支持 4 万个神经元、1000 多万个突触。

每颗芯片支持 625 万个神经元。

由单芯片组成的神经网络（625 万个神经元网络）。

将基于脉冲神经网络（SNN）的类脑算法和基于人工神经网路（ANN）的深度学习算法集成到一颗芯片上。

④ 浙江大学+之江实验室， 达尔文-1 芯片，2015 年发布

达尔文-2 芯片，2019 年发布

达尔文-3 芯片， 尚在研制中

每颗芯片 576 个内核，每个内核支持 256 个神经元，6 万个突触。

每颗芯片支持 15 万个神经元。

达尔文-2 芯片集成：792 颗芯片支持 1.2 亿个脉冲神经元、300 亿个突触，相当于果蝇大脑神经元数量规模。

⑤其他单位类脑芯片

曼彻斯特大学： SpinNaker 2018.11.2 发布。

其芯片集成： 支持 100 万个内核、10 亿神经元，相当于小鼠大脑神经元规模。

555. 类脑计算（神经拟态计算）与传统硅基计算、传递、运行模式

①前者的信息源是神经电脉冲信号和化学信号，后者为数字信号

②前者编码方式采用稀疏脉冲时序编码机制，后者由数字源代码变换为 0,1 的机器码

③前者信息传递方式通过模仿人脑自然神经元+突触组成的脉冲神经网络和运行方式模型，对神经电脉冲进行信息传递，后者采用传统网络权重连接+激活方式，对机器码进行信息传递

④前者计算（神经元）和存储（突触）是一体化的、融合在一起的，后者计算（处理单元）和存储（存储单元）是分离的

⑤前者存在三维广泛连通性，后者无法模拟三维连通，受限于二位连接

⑥前者是基于脉冲的事件驱动型的随机计算，后者为了构建确定性计算采用晶体管间布尔代数电路开关

⑦前者类脑是模拟生物学上自主低能耗类型（计算/处理、传输），实行节能型，后者能耗大

⑧前者模仿人脑的大规模进行通信方式（结构），将数十亿信息同时送到数千个不同目的地，后者也可采用并行通信方式，但传送信息为通过标准网络（从A点到B点），发送大量信息进行通信

⑨前者其运行方式符合类脑计算（神经拟态计算）系统的运行方式(或神经拟态计算架构)，后者其运行方式符合冯·诺依曼计算架构。

556. 类脑计算完备性

这是清华大学计算机系张悠慧教授团队和精仪系施路平教授团队合作撰写的一篇论文（论文作者为：Youhui Zhang, Peng Qu, Yu Ji, Weihao Zhang, Guangrong Gao, Guanrui Wang, Sen Song, Guoqi Li, Wenguang Chen, Weimin Zheng, Feng Chen, Jing Pei, Rong Zhao, Mingguo Zhao & Luping Shi ）。论文题目：《A system hierarchy for brain-inspired computing》，是清华大学今年第三次在《Nature》杂志上发表的文章（2020.10.14）:

本文摘要

提出神经拟态完整性

大脑启发计算目前缺乏一个简单而健全的系统层次来支持整体开发。因此，神经拟态软件和硬件之间没有清晰完整的接口。由于许多灵感来自大脑的芯片不是为传统的通用计算而设计的，它们中很少提供传统的指令集，因此不清楚它们是否是图灵完成的。而图灵完备性是传统编译的可行性基础，要求程序的表达和转换是等价的。所以本文提出神经拟态完备性，这是一种更适用于大脑启发计算的完备性的更广泛定义。它放宽了神经拟态硬件的完备性要求，提高了不同硬件和软件设计之间的兼容性，并通过引入一个新的维度—近似粒度来扩大设计空间。

提出一种系统层次结构

神经拟态计算与传统计算的区别还在于：它使用同步计算和存储，使用基于spikes (spiking 神经网络的特征)的事件驱动计算，并在高并行方面具有更大的潜力。这些差异使得传统的计算机层次结构难以直观地描述类脑应用程序。因此提出了一种具有高通用性的类脑计算系统层次结构。这个层次结构有三个层次：软件、硬件和编译。

软件层次，提出了一个统一的、通用的软件抽象模型——编程操作符图(programming operator graph, POG)——以适应各种大脑启发的算法和模型设计。该模型集成了存储和处理。它描述了什么是大脑激发程序，并定义了它是如

何执行的。由于 POG 是图灵完成的，它最大程度地支持各种应用程序、编程语言和框架。

硬件层次，设计了抽象神经拟态体系结构(ANA)。

编译层次，是将程序转换为硬件支持的等效形式的中间层。为了实现可行性，提出了一套被主流的脑激发芯片广泛支持的基本硬件执行原语，证明了配备这套硬件的神经形态是完整的。最终通过实验验证了神经形态完备性引入的系统设计层次的优化效果。

相关实验部分

第一个应用实验是一种用于自行车驾驶和跟踪的人工神经网络模型。

第二个应用实验是用于鸟群模拟的 boids 模型。

第三个应用实验是 QR 分解（非线性计算的数学算法）。

神经拟态计算从生物大脑中汲取灵感，为计算技术和体系结构提供了推动下一波计算机工程发展的潜力。这种受大脑启发的计算也为人工智能的发展提供了有前途的平台。传统的计算机系统具有围绕图灵完备和冯·诺伊曼体系结构建立的完善的计算机层次结构，而与传统的计算机系统不同，目前尚无广义的系统层次结构或对类脑计算的完整性的理解。这会影响软件和硬件之间的兼容性，从而阻碍类脑计算的开发效率。

清华大学该团队提出了“类脑计算完备性”，它放宽了对硬件完整性的要求，并提出了相应的系统层次结构，其中包括图灵完备的软件抽象模型和通用的抽象神

经形态架构。使用这种层次结构，可以将各种程序描述为统一的表示形式，并转换为任何神经形态完整硬件上的等效可执行文件。也就是说，它可以确保编程语言的可移植性，硬件完整性和编译可行性。该团队实现了一系列工具链软件用以支持在各种典型的硬件平台上执行不同类型的程序，进而证明了系统层次结构的优势。

希望可以使类脑计算系统的各个方面实现高效且兼容的进展，从而促进包括人工智能在内的各种应用程序的开发。

论文链接：https://www.nature.com/articles/s41586-020-2782-y

557. 解决最后一公里短板，提升语义网络内涵，使之具有实现可解释人工智能（或实现认知智能）的能力，尚待努力！

在跟贴 263、521 中，谈到提升语义网络内涵的解决之道，要坚持数据（第二代人工智能）、知识（第一代人工智能）融合统一的双驱动。

在谈到语义网络建设中未及攻克的难关时，跟贴 530 指出，不少知识是目前语义网络还不能概括的，如常识，常识是难以定义、表达、表征的，目前的大规模语义网络尚不包括常识，如常识外，还有背景知识、专业知识、专家经验、隐性知识等，也不能被大规模语义网络所概括。

跟贴 457 也指出，常识是无法穷尽不成文的规则，是一种广泛可重复使用的背景知识。建立常识库，以此实现自动化常识推理第一步，但做起来难度之大到难以想像！

在跟贴 457 中也谈到：人工智能要变得像人一样聪明，常识推理能力是必备的，背景知识在学习和训练时是不可或缺的，对于常识、专业知识、专家经验，机器是很难识别的。机器缺去常识推理，何时到了破局的时候？！

在 1-551 条跟贴中，不少专家致力于研发大规模语义网络（知识图谱）中可理解、可解释的内函。以跟贴 457 为例：

OpenAI 于 2019 年公布 GPT 一 2：具有 15 亿参数的通用语言模型，这在语言模仿上有较大进展，但它还是缺乏基本常识。

跟贴 457 推荐华盛顿大学叶锦才（YejinchoYejinchoj）研发团队关于常识推理攻关研究进展。他们提出了自动知识图谱构建模型 COMET，融合了 GOFAI 式的符号推理和深度学习（知识和数据双驱动）两种人工智能方法。

近年来，大规模语义网络（将人类语言转化为机器可理解的内容）有很大进展，同时建设大型常识库也有不少探索，如果有人要查阅这方面的资料，推荐查阅跟贴 1-556。

558. ISO 的可解释 AI 标准项目

2020 年是国际标准组织 ISO/IEC JTC1 成立人工智能标准分委员会 SC42 的第三年。在不久结束的第六次年会上，SC42 批准将新标准项目提案《机器学习模型和人工智能系统的可解释性之目标和方法》提交 SC42 的全体成员国投票表决。这一项目旨在本文档描述可用于实现 ML 模型和 AI 系统的行为、输出和结果方面的不同利益相关者的可解释性目标的方式方法。所谓利益相关涉及学术界、产业界、

政策制定者和最终用户等。不出意外，此项目将最快明年春天开始，预计 2020 年中完成。

559. 机制主义通用智能理论

（钟义信教授来信）

陆总：您好！您寄来的《国内外 AI 跟贴留言》第六集收到了，非常感谢！

向您报告：我们原先发现的“普适性智能生成机理”不但适用于人工智能，而且适用于人类智能。因此，“机制主义通用人工智能理论”可以推广成为“机制主义通用智能理论”。这里的关键就在于“研究范式的革命”。

这太有意思了！

如果您有兴趣和时间，我可以给您做一个具体的汇报。

钟义信

PS: 559 号附件：人工智能范式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（见跟帖 584 条之后）

560. MIT 评论：机器学习模型构建中的缺陷

我们训练的人工智能的方式根本上是有缺陷的

我们今天使用的大多数机器学习模型的构建过程无法判断它们在现实世界中是否有效，这是一个问题。

https://www.technologyreview.com/2020/11/18/1012234

561. 全球最大的人工智能研发组织 LFAI 基金会执行董事 Ibrahim Haddad 应邀在《第 15 届开源中国开源世界高峰论坛（线上会议）》上作报告，介绍 LFAI 组织由上千家公司、10 所大学组成（迄今为止）。Ibrahim 答复陆主席询问哪 10 所大学？如下：①治亚理工学院，②哥伦比亚大学，③斯坦福大学，④宾夕法尼亚大学，⑤纽约大学，⑥麻省理工学院（大学），⑦哈佛大学，⑧牛津大学，⑨佛罗里达州立大学，⑩哥本哈根大学。陆主席提出似乎还缺卡耐基梅隆大学，还需发展中国的一些大学，Ibrahim 表示正在改进，正在与中国的鹏城实验室洽谈合作。

562. 清华、北大教授同台激辩：脑科学是否真能启发人工智能？

教授们认为，脑科学是智能科学一个重要的研究方向，但不是做人工智能的前提，它与人工智能应该是一个相辅相成的过程。

目前认知神经科学取得极大进步，对现在的人工智能的端对端学习与强化有诸多启发，但认知科学的进展对下一代人工智能是否有帮助？脑科学是否能真正启发人工智能？都是有待探索的问题。

有的教授提出，能否构造不同于人脑的认知智能系统？

进化不是线性的，是一棵进化树。不同生物有不同的神经元，但所有生物（高级的、低级的）对世界都有非常好的适应。所以无论其神经元多少，都能做到通用智能。

不同生物的神经突触连接是不一样的，据此如果完全按照生物系统以物理上的神

经结构去做智能，真的能够得到普遍性的规律吗？

人工智能有很多途径：主流的是由数理逻辑和专家系统去做符号主义；另外一个是连接主义，构造一个大的神经网络；还有行为主义，主要是从控制论去做。人脑是一个近乎完美的通用智能系统，所以当连接主义出现问题和瓶颈时，人脑的存在可提供一个最终的方向和信念。

计算机里的任何一个问题，并没有最优算法，都是多个最优算法同时存在，所以实现人工智能也应该有多条路径。

如果我们想从人类智能或生物智能借鉴一些东西帮助人工智能发展，最核心的是理解人脑，这需要去深入挖掘。

现在的人工智能是通过计算机、机器实现的，其根本的是要适应机器而不是人脑。

对脑科学或人工智能最重要的应该是存在性的东西，智能是存在的，但是脑科学目前并没有提供一种途径实现人工智能。

下一代人工智能最需要去关注什么？

哪些人类或生物体的认知功能是下一代人工智能最需要去关注或借鉴的？

当下生物体智能和人工智能有哪些差异？

大脑与人工智能在结构上不同，结构不同实际是编码集不同。

神经系统还有 Conference（来自先验知识），通过神经元反应的随机性把 Conference 编码在一起，机器学习中的 Conference 是通过大数据做出来的，现在的问题是，神经系统发生的机制和机器学习是否一样？

在有限资源下重要的是学习、记忆及遗忘。知识是人区别于动物最本质的东西。脑神经是比较简单的物理系统，把它变成向量在逻辑上是通的，但是向量比较精确，脑能否处理这么复杂的事？

现在的知识图谱是三元组，三元组做推理是没有问题，但脑子里是不是三元组？大脑的神经元是多样性的，但目前神经网络里的神经元都长一个样子，去做这样一个网络模型很困难。

大脑是复杂的网络动力学模型，引入神经元里的动力学系统后，能做记忆和联想，如何把这种类似于神经动力学的系统反映在一个计算系统上是一个比较关键的问题。

大脑里有很多可以借鉴的东西，如稀疏编码、注意力机制，能把多层次多精度的记忆和联想在计算系统里做实现。而现有计算机体系存算分离，计算系统中所有记忆都一样。如神经网络存的全是权重，没有把记忆分层次粒度。因此如何把记忆跟计算系统及决策联系起来也是一个关键问题。

人工智能如何与各学科联合发展？

563. 创建结构数学研发通用 AI

大脑是模拟的，机器（计算机）是数字的，深度学习将数字输入大脑，大脑是不识别的，而大脑中的思维逻辑传到机器，机器也是不识别的。所以采用深度学习算法技术难以做到脑机交互。

研发下一代通用人工智能，从计算理论角度讲，巴贝奇与图灵计算模式均不适用，

需要创建新的计算理论基础，统一时空计算理论，并与脑科学统一；目前的障碍在于数学描述方法，要创建一门新的数学——结构数学，以解决结构化多因果描述全息结构计算模型。

564. 设计 2—16—256 进制类脑计算机

现在已经实现设计 16 进制的类脑计算机。这时不仅计算速度、计算能力能呈指数级增长，关键是计算性能更符合人脑功能机制。

565. 华为自动驾驶异军突起！日前华为参加一项比赛，华为自动驾驶方案与诺亚方舟实验室参赛队伍的表现极为突出，力压西门子等参赛队伍成功夺冠！

566. 中国量子计算机原型机“九章”研制成功

中国科大潘建伟团队（与中科院上海微系统所、国家并行计算机工程技术研究中心合作）于今年 12 月 4 日发布成功构建了 76 个光子的量子计算机原型机“九章”，实现具有实用前景的“高斯玻色取样”任务快速求解，该机处理这项特种任务运算 200 秒，比今年最快的超级计算机“富岳”快 100 万亿倍（“九章”1 分钟完成任务，起算需 1 亿年），“九章”比谷歌去年发布的构建 53 个超导比特量子计算机原型机“悬铃木”运算速度快 1 万倍。

567. 构建“类脑智能”（或“仿脑”）不一是完全“克隆人脑”（或“复制类人”）有人提出下一阶段人工智能即通用人工智能（强人工智能），其目标是构建“类脑智能”，不应该是构建一个与人脑完全一样的东西，是一个模拟人脑的智能，

不是完全克隆人脑的“类人智能”!

568. 刘江老师致电 COPU 秘书处谈人工智能

人工智能最近比较值得关注的是大规模预训练模型，以 BERT 和 GPT-3 为代表。其他没啥特别大的进展。我现在智源人工智能研究院兼职副院长，信息很多，陆老有什么 AI 方面的问题可以随时发我。

LFAI 说是最大 AI 研发组织，这个说法不严谨。AI 研发重镇还是 Google、微软和 Facebook 等美国大互联网公司，还有 CMU、斯坦福、MIT 等顶级高校，主要围绕顶级学术会议运作。LFAI 主要还是传统 IT 公司在玩（IBM、华为、爱立信等），学校也不是最好的，所以在 AI 圈子里影响并不大。

请代转给陆老。

陆主席答复如下:

转告刘江，他说的两点：大规模预训练模型，2018 年谷歌发布的 BERT、2020 年 OpenAI 发布的 GPT 一 3 只是对大规模语义网络进行微调达到精确掌握自然语言的目的，这在我们发布的人工智能国内外跟贴中曾作过大量的介绍、研究与评论，现在的问题是欲以大规模语义网络破解认知智能（或建立可解释、可理解的人工智能还差最后一公里），希望刘江好好研读跟贴，在此基础上我们可展开进一步讨论。关于 LFAI 不久前才创建，我与该基金会执行董事 Ibrahim 交谈过多次，Google、微软、Facebook、CMU、斯坦福、MIT 均是其主要成员，人工智能项目孵

化器的主持者，IBM 也是，为此还请刘江能认真阅读 600 多条（迄今）跟贴。在这里我还是要赞赏高文院士，LFAI 告知我，鹏城实验室正在与他们洽谈合作，我当然可助力促进!

569.《通用智能理论》创新思想梳理

陆总：这是我的与众不同的观念和方法。请您批评!

———钟义信　北京邮电大学人工智能学院

《通用智能创新思考梳理》

一、怎样发现问题？—看本质

（1）

人工智能存在的主要问题之一

是学科的整体被肢解

（鼎足三分，无法形成整体理论）

而不是通用性够不够、移植性好不好？

（2）

人工智能存在的主要问题之二

是它的“智能”被置空

（没有理解能力、结果不可解释）

而不是理解力好不好、可解释性强不强

（3）

整体被肢解：归因于“分而治之”方法论

智能被置空：归因于“纯粹形式化”方法论

总之，归因于研究范式

不是仅靠算法、算力、数据的改善就能解决

也不是把结构类脑 / 功能类脑结合就能解决

或者把“数据驱动”与“知识驱动”相结合就可以

二、怎样寻找解决问题的方法—循规律

（4）

只有突破自然科学与社会科学哲学之间的藩篱

才能看清人工智能问题的根源在范式

局限于自然科学范畴内，只能看到算法、算力、数据

只能看到脑结构、脑功能

（5）

突破藩篱之后就可以看出

驾驭学科研究的最高引领力量是学科的研究范式

科学观回答：学科是什么？

方法论回答：学科怎么做？

两者共同决定了学科的规范定义

否则，学科定义就有随意性

（6）

范式，是统领学科研究的龙头

首先自下而上探寻范式

然后依照范式（学科定义）自上而下建构学科：

学科定义（定义）

学科框架（定位）

学科规格（定格）

学科理论（定论）

人们往往只看到“学科理论”，看不到它的根基

（7）

调研发现

人工智能问题的根源是：范式的张冠李戴！

（用传统学科的范式指导人工智能研究）

所以，人工智能的根本问题是：范式革命

其他研究都解决不了根本问题

三、怎样解决问题—把范式革命贯彻到底

（8）

为此，需要建立信息科学的研究范式

科学观：主客互动信息过程，而非物质客体

方法论：信息生态论，而非分而治之 / 纯粹形式化

（9）

人工智能的全局模型是

主体驾驭下的主客互动信息过程

而不是孤立的脑

（10）

人工智能的研究路径是

建立普适性的智能生成机制

而不是结构、功能、行为的分立路径

（11）

人工智能的学科结构是

原型科学、本体科学、基础科学、技术科学的交叉综合

而不是计算机科学的分支

（12）

人工智能的基础科学是

泛逻辑理论和因素空间数学理论

而不是传统的概率论和数理逻辑

（13）

人工智能的基本概念是

形式内容价值三位一体的全信息、全知识、全智能

而不是纯粹形式化数据、形式化知识、形式化智能

（14）

人工智能的基本原理是

信息转换与智能创生定律

它与

质量转换与物质不灭定律

能量转换与能量守恒定律

构成物质、能量、信息的三大定律

（15）

普适性的智能生成机理

既适合于人工智能，也适合于自然智能

但实现机理的方式各有特色

（16）

基于以上的突破、发现和建构

创立了《通用智能理论》

不是“不存在通用智能理论”

570. COPU 谈人工智能专题会议

——COPU 开源联盟秘书处

12 月 1 日（周二）陆主席召开 COPU 专题会议，讨论人工智能国内外 551 条跟帖评论:

跟帖涵盖全球人工智能研发前沿，有 6 大部分，

1) 不可解释的机器学习/深度学习支持今天人工智能的繁荣，说它已无发展潜力、已近天花板是不妥的;

2) 打破机器学习的黑盒子研发可解释的人工智能，有所突破;

3) 基于异步脉冲神经网络的神经拟态计算系统，已有亮点;

4) 从知识工程出发，依托大规模语义网络（知识图谱）的支持，用以破解认知智能解决方案，还差最后一公里;

5) 脑机接口的理论和实践，目前国内外已有几十例试点;

6) 钟义信教授、张钹院士对不同学派提出的人工智能发展路径提出质疑，但钟、张所提的发展模式还是概念。此处还报导了清华、北大教授激辩：脑科学是否真能启发人工智能？！

现在看来，目前人工智能国内外跟帖评论热闹非常！朋友，如你有兴趣，不妨参加进来（先看后评）。

571. 李德毅院士关于通用人工智能十问

钟义信教授、张钹院士分别提出要构建作为下一代人工智能的通用人工智能（或

强人工智能)。

通用人工智能即下一代人工智能，需要全新的理论基础，当前无论从哲学层面讲，或从科学层面讲，尚不完备!

今天人工智能的目标不是构建与人脑完全一样的东西，也并非全知全能，而是模拟人脑智能（即构建类脑智能)。

如果没有人工智能全新的理论突破，没有信息处理机制全新范式，李院士的“十问”在现有人工智能理论框架内将难以回答，而如不能回答这些问题，那么通用人工智能也就很难实现。

下面是李德毅院士关于通用人工头智能的十问:

一、意识、情感、智慧和智能，它们是包含关系还是关联关系？是智慧里面有智能，还是智能里面有智慧？大凡意识、情感都是内省的、自知的、排他的，怎么可以用他人的、人工的来代替呢？所以非生命体不可能有意识?

二、如何理解通用智能？我们应该不应该把通用智能理解为“全知全能”或者单向超强智能？尽管今天的计算机已经可以解决很多复杂的、专门的智力问题（如围棋智能)，我们仍常常觉得它们缺乏人类思维的某些本质特征。这里的差别主要不是在算法、算力、数据量方面，不是在速度和容量方面，而是在智能的一般性、通用性、普遍性、灵活性、缺省性、容错性、可习得性、不确定性、适应性、常识性、开放性、创造性、自主性等方面。遗憾的是发展 60 多年的人工智能没有能够更靠近人的原始的智能。

三、目前所有的人工智能的成就都是在计算机上表现出来，是基于冯架构的计算机智能或者计算智能，人工智能是计算机的一个应用而已。而人脑不是冯诺依曼架构的，存在不存在宏观上更类似脑的非冯诺依曼架构呢？例如，对人的智能而言，记忆力是真正的智力，超强记忆力就是超强智能，记忆比计算机重要，记忆的提取要比复杂的推理快得多，非冯架构如何在结构上体现人脑的不同记忆区和记忆力呢？如何体现环境和知识的双驱动？

四、非生命体不会有七情六欲，机器人是非生命体，还会有学习的原动力吗？如果没有学习的原动力，没有接受教育的自发性，还会有学习的目标吗？目标从哪产生？机器人能否自己提出问题？

五、人的注意力选择源于记忆，源于记忆的偏好依附性，偏好如何产生的？偏好依附是否只能与交互认知的频度和时间的远近相关？人的偏好依附不是这样的，人的恐惧性以及满足感会让一些发生频度很低、或者很久远的事记忆特别深刻。

六、自然语言是人类思维活动的载体，如果自然语言是第一语言，数学语言是第二语言，计算机语言是第三语言，后一个比前一个常常更严格，后一个比前一个常常更狭义，根据哥德尔不完全定理，数学自身难以完全自洽。数学的形式化要借助于自然语言，计算机语言的形式化要借助于数学语言。因此，人工智能怎么可以反过来要用数学语言或者计算机语言去形式化人类的自然语言呢？

七、人脑是个小宇宙，其中的智能是多情境、多公理兼容并包的，在不同情境里

有不同应对，不完全收敛，不完全自恰，不整体统一，不存在非公理的统一的数学推理，当然也不必一定要脑裂。

八、一个机器或者系统是否有智能，不在于某一个时刻它能解决什么实际的智力问题，而在于它有没有学习的能力？智能，即提供的问题解决方案，是否依赖于有限的认知资源？是否需要进一步交互认知？是否可以有选项？是否可以进化和成长？这才是最重要的。

九、在一个非冯诺依曼架构的机器人脑中，组成记忆、交互和计算的基本元件最少有哪几种？各元件中的信息的产生机制与存在形式是什么样的？他们之间的信息传递机制是什么样的？

十、通用智能后天的习得靠教育，智能植根于教育，文明是智能的生态。

572. 谷歌最新人工智能 AlphaFold-2（开始进军基因医疗科学），基于机器学习解决蛋白质折叠问题。谷歌最新人工智能 AlphaFold-2，根据氨基酸序列预测了生命基本分子一蛋白质的三维结构，将其人工智能转向了人类科学中最棘手的领域—基因医疗科学！

2018 年 AlphaFold（在人工智能国内外跟贴 66 条、67 条、68 条、69 条、70 条、213 条中均有介绍）便已经取得了里程碑的突破（用数千种已知蛋白质训练深度神经网络，成功地根据基因序列预测了生命基本分子一蛋白质的三维结构）。但是仍然没有完全解决蛋白质折叠问题。这次 AlphaFold-2 直接斩获 92.4 的准确

性高分，其误差基本不超过一个原子的大小。

“蛋白质折叠”是一种令人难以置信的分子折纸形式。所有生物都是由蛋白质构成的，蛋白质的结构决定了它的功能。一旦蛋白质折叠错误，就会导致糖尿病、岶金森症和阿尔茨海默病等疾病。

预测蛋白质折叠结构的能力意义重大，AlphaFold 的出现，意味着今后要对蛋白质结构进行高效、简便但精准的预测，仅需初步的试验数据即可。拥有这些蛋白结构的助力，疾病、演化等领域的研究将得到强大的推动。

为了开发 AlphaFold，谷歌用大量已知蛋白质训练神经网络，今年参加 CASP 的 AlphaFold，训练数据集囊括了大约 17 万个已知的蛋白三维结构，利用 128 个 TPUv3 核心，AlphaFold 在训练几周后就达到了参赛水平。

今天，谷歌 AlphaFold 成功预测蛋白质的三维结构表明，当人工智能与基因科学相结合，人类将进入一个风高浪急的新时代。

这一巨大突破，直接引爆了全网。《Nature》杂志评论：这将改变一切！

573. 读李德毅院士十问后

——读者 SQ1204

读了李院士的十问，关于研发通用人工智能即下一代强人工智能的思路，他似乎也倾向于钟院长、张院士所提的发展概念（在此概念下未来的具体模式李、钟、张未必相同）。

他们三人的目标都是研发类脑智能（通用人工智能），他们认为通用人工智能需

要全新的理论基础，目前无论从哲学层面上讲，或从科学层面上讲，均不完备！目前人工智能的成就都是基于冯-诺伊曼架构的计算智能，缺乏人类思维的某些本质特征，这里主要差别不在于算法、算力、数据量方面。人工智能迄今已产生的若干理论与技术流派，如神经网络、知识工程、行为机制和现场学派都从不同侧面模拟人脑智能，已显露出其片面性和局限性，在智能突破性方面进展有限。上述三位老师是否如你们所预言的（？）对今天不同学派提出的人工智能发展途径要统统排斥、抛弃？！可是据已发布的500多条国内外跟贴的批露，其中一些所谓片面的发展途径已实现了可解释的人工智能，在进军类脑智能方面似乎也已见端倪，在你们还在作空谈之时，他们似乎更为求实！

574，可解释的信用评分模型（为信贷建立可解释的信用评分模型）

Malta 大学人工智能部 LaraMarieDemajo，ＶinceＶella，AlexieiDingli，2020.12.4

随着人工智能和金融技术的发展，信用评分模型已经引起了学术界广泛关注。信用评分可以帮助金融专家更好地决定是否接受信贷申请。最近一些法规，如《一般数据保护条例》（GDPR）和《平等信贷机会法（ECOA），都增加了对模型可解释性的要求，以保证算法决策的可理解性和一致性。论文作者提出了一个既准确又可解释的信用评分模型。该模型的流程是：首先对数据进行预处理，然后使用XGBOOST模型对数据实例进行分类，最后使用三种XAI方法对分类器进行扩展，提供一个全方位的解释框架。

数据预处理过程：数据清洗→特征生成/选择→数据集划分→标准化→交叉验证→平衡数据。

不同的人在不同情况下需要不同的解释，而单一的 XAI 方法不足以提供所有的解释。因此作者提出了一个可解释的信用评分模型，为各种角色产生一种解释。在信用评分模型中有三种不同的角色：

①信贷员，喜欢基于实例的局部解释。该解释提供了对单个实例预测的局部解释。信贷员更倾向于这样的解释，因为他们需要知道模型给出的预测结果是否合理。

②被拒绝的贷款申请人，喜欢基于特征的局部解释。这项解释是对某一特定预测结果的解释，说明模型是如何做出该预测的以及这样预测的原因。贷款申请人更倾向于这样的解释，因为他们最关心的是为什么他们的贷款申请被拒绝。

③监管者或数据科学家，喜欢模型全局解释。这项解释是对模型整体工作方式的理解，解释了模型在做预测时背后使用的逻辑推理。监管机构和管理层通常更倾向于这样的解释，因为他们主要关注的是对信用评分模型的全局理解，而不是对每种情况的个别解释。

列出为各种角色提供解释的 XAI 方法如下：

Explanation Type	**XAI Method**	**Explanation Form**
Global	SHAP ＋ GIRP	Decision Tree/IF-THENrules
Local feature-based	Anchors	DNF rule
Local instance-based	ProtoDash	Prototypicalinstances

最后，作者进行实验，检验了模型的正确性、有效性、易理解性、细节充分性和可信度。

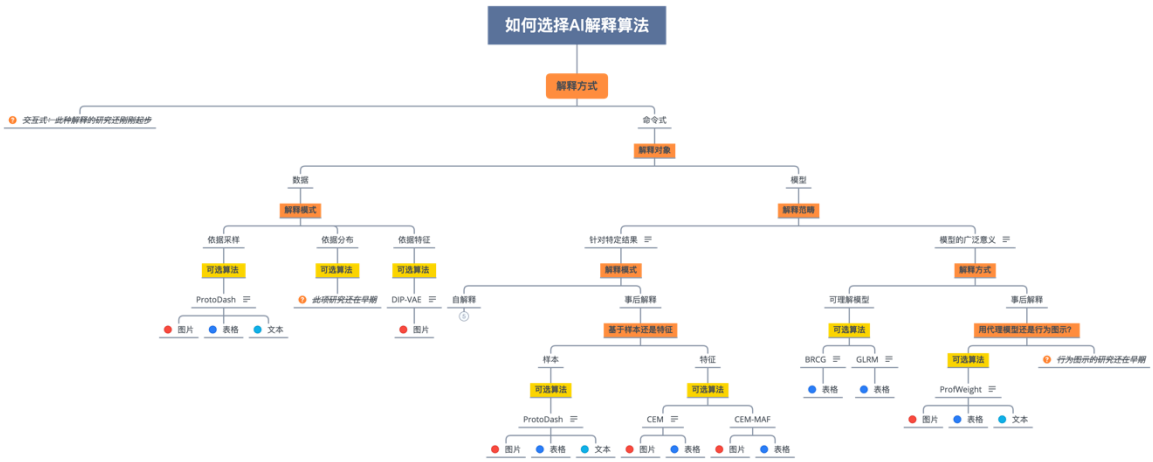
575. 可解释 AI 的上手实践

——IBM 田忠博士摘自 IBMAI 研究院

人工智能（AI）有意无意间已经成为我们生活的一部份。如何安心、放心、信心使用 AI 成为业界研究和实践的重点。学术界的研究相当活跃，每年都有若干专门的学术会议，如 WHI2020 （Workshop on Human Interpretability in Machine Learning)。2019 年,IBM 研究院多年关于可解释 AI 的研究成果开源,合称 AIX360,并捐给 Linux 人工智能基金会（LF AI)。 截止 2020 年初，AIX360 包含的算法，其中 8 个来自 IBM 研究院的科研成果，2 个是业界的流行算法，有关的代码、文档、演示可在 http://aix360.mybluemix.net 获得。和 AIX360 相辅相成，致力于可信赖 AI 的其它项目还有致力于公平性的 AIF 360、致力于健壮性的 ART 360、致力于真实性的 Factsheet 360。

可解释 AI 的意义、目的、方法因人而异。对于 AI 系统开发者、数据科学家、项目经理而言，其目的多半是如何提高系统效率；对于 AI 系统使用者，如医生、律师、银行贷款经理、考官，则是需要对 AI 系统做出的推荐需要的是信心、放心、安心；对于主管当局，如欧盟委员会、纽约市政府、中国银保监会，他们主要关心的是如何确保 AI 的系统性的公平性；而对于最终受影响的用户，如病人、诉讼对象、贷款申请人、教师，他们需要的是能够理解影响结论的主次要因素，从而未来能够有所作为。

那么如何按需选择合适的解释算法呢？下面的树形结构可供参考。



我来看一个示例，一家银行使用了 AI 系统帮助基于公开可获得的 FICO HELOC Dataset 真实数据来辅助决策是否批准一项贷款申请。对于建设系统的数据科学家、银行的贷款经理以及贷款申请者对系统的解释性有不同的需求，因而选择了不同的算法获得洞察。

对于建设本 AI 系统的数据科学家来说，重中之重是向银行主管以容易理解的方式（比如一组简明规则）解释本系统的工作效果。为此，他需要系统执行一个命令来获得对决策模型的普适意义的解释模型。依照上面的选择路径，他因此选择了 BRCG 算法以生成一组布尔规则表，使用 GLRM 算法生成逻辑规则回归模型。

有了这个决定，他按如下步骤开展：加载整理数据、运行算法、显示结论。

加载整理数据

	8960	8403	1949	4886	4998
ExternalRiskEstimate	64.0	57.0	59.0	65.0	65.0
MSinceOldestTradeOpen	175.0	47.0	168.0	228.0	117.0
MSinceMostRecentTradeOpen	6.0	9.0	3.0	5.0	7.0
AverageMInFile	97.0	35.0	38.0	69.0	48.0
NumSatisfactoryTrades	29.0	5.0	21.0	24.0	7.0
NumTrades60Ever2DerogPubRec	9.0	1.0	0.0	3.0	1.0
NumTrades90Ever2DerogPubRec	9.0	0.0	0.0	2.0	1.0
PercentTradesNeverDelq	63.0	50.0	100.0	85.0	78.0
MSinceMostRecentDelq	2.0	16.0	NaN	3.0	36.0
MaxDelq2PublicRecLast12M	4.0	6.0	7.0	0.0	6.0
MaxDelqEver	4.0	5.0	8.0	2.0	4.0
NumTotalTrades	41.0	10.0	21.0	27.0	9.0
NumTradesOpeninLast12M	1.0	1.0	12.0	1.0	2.0
PercentInstallTrades	63.0	30.0	38.0	31.0	56.0
MSinceMostRecentInqexcl7days	0.0	0.0	0.0	7.0	7.0
NumInqLast6M	1.0	2.0	1.0	0.0	0.0
NumInqLast6Mexcl7days	1.0	2.0	1.0	0.0	0.0
NetFractionRevolvingBurden	16.0	66.0	85.0	13.0	54.0
NetFractionInstallBurden	94.0	70.0	90.0	66.0	69.0
NumRevolvingTradesWBalance	1.0	2.0	10.0	3.0	2.0
NumInstallTradesWBalance	1.0	2.0	5.0	2.0	3.0
NumBank2NatlTradesWHighUtilization	NaN	0.0	4.0	0.0	1.0
PercentTradesWBalance	50.0	57.0	94.0	46.0	83.0

运行算法

1）BRCG

```
# Instantiate BRCG with small complexity penalty and large beam search width
from aix360.algorithms.rbm import BooleanRuleCG
br = BooleanRuleCG(lambda0=1e-3, lambda1=1e-3, CNF=True)

# Train, print, and evaluate model
br.fit(dfTrain, yTrain)
from sklearn.metrics import accuracy_score
print('Training accuracy:', accuracy_score(yTrain, br.predict(dfTrain)))
print('Test accuracy:', accuracy_score(yTest, br.predict(dfTest)))
print('Predict Y=0 if ANY of the following rules are satisfied, otherwise Y=1:')
print(br.explain()['rules'])
```

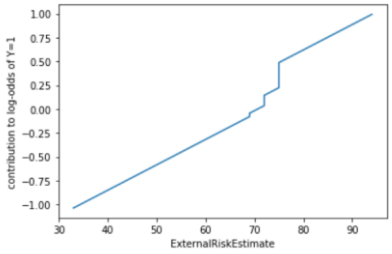
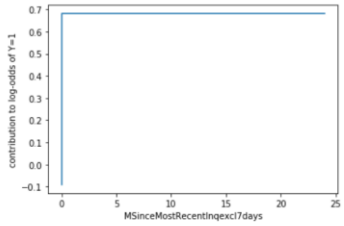
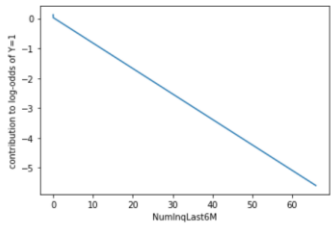
```
Learning CNF rule with complexity parameters lambda0=0.001, lambda1=0.001
Initial LP solved
Iteration: 1, Objective: 0.2895
Iteration: 2, Objective: 0.2895
Iteration: 3, Objective: 0.2895
Iteration: 4, Objective: 0.2895
Iteration: 5, Objective: 0.2864
Iteration: 6, Objective: 0.2864
Iteration: 7, Objective: 0.2864
Training accuracy: 0.7195731460218 83
Test accuracy: 0.696515397082658
Predict Y=0 if ANY of the following rules are satisfied, otherwise Y=1:
```

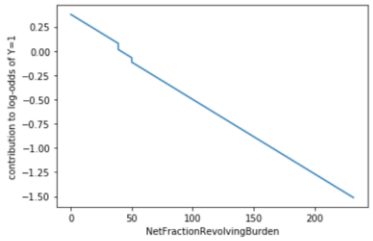
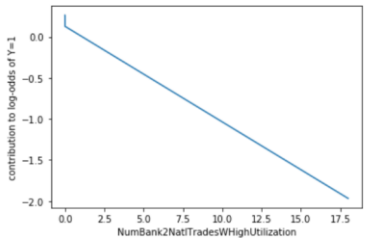
2）LogRR

```
# Instantiate LRR with good complexity penalties and numerical features
from aix360.algorithms.rbm import LogisticRuleRegression
lrr = LogisticRuleRegression(lambda0=0.005, lambda1=0.001, useOrd=True)

# Train, print, and evaluate model
lrr.fit(dfTrain, yTrain, dfTrainStd)
print('Training accuracy:', accuracy_score(yTrain, lrr.predict(dfTrain, dfTrainStd)))
print('Test accuracy:', accuracy_score(yTest, lrr.predict(dfTest, dfTestStd)))
print('Probability of Y=1 is predicted as logistic(z) = 1 / (1 + exp(-z))')
print('where z is a linear combination of the following rules/numerical features:')
lrr.explain()
```

```
Training accuracy: 0.742536809401594
Test accuracy: 0.7260940032414911
Probability of Y=1 is predicted as logistic(z) = 1 / (1 + exp(-z))
where z is a linear combination of the following rules/numerical features:
```

	rule/numerical feature	coefficient
0	(intercept)	-0.0686341
1	MSinceMostRecentInqexcl7days > 0.00	0.680261
2	ExternalRiskEstimate	0.654248
3	NetFractionRevolvingBurden	-0.553965
4	NumSatisfactoryTrades	0.551654
5	NumInqLast6M	-0.463226
6	NumBank2NatlTradesWHighUtilization	-0.448331
7	AverageMInFile <= 52.00	-0.43436
8	NumRevolvingTradesWBalance <= 5.00	0.42154
9	MaxDelq2PublicRecLast12M <= 5.00	-0.418142
10	PercentInstallTrades > 50.00	-0.317566
11	NumSatisfactoryTrades <= 12.00	-0.312471
12	MSinceMostRecentDelq <= 21.00	-0.301566
13	PercentTradesNeverDelq <= 95.00	-0.273924
14	ExternalRiskEstimate > 75.00	0.263437
15	AverageMInFile <= 84.00	-0.182118
16	PercentTradesNeverDelq	0.166518
17	AverageMInFile	0.15069
18	PercentInstallTrades > 42.00	-0.148802
19	NumBank2NatlTradesWHighUtilization <= 0.00	0.135396
20	MSinceOldestTradeOpen <= 122.00	-0.132409
21	PercentTradesNeverDelq <= 91.00	-0.11771
22	NumSatisfactoryTrades <= 17.00	-0.11022
23	ExternalRiskEstimate > 72.00	0.107613

（图形）显示结论，如以 GAM 图示 LogRR 的结论	外部风险预估
	ExternalRiskEstimate
	As expected from the BRCG Boolean rule above, 'ExternalRiskEstimate' is an important feature positively correlated with good credit risk. The jumps in the plot indicate that applicants with above average 'ExternalRiskEstimate' (the mean is 72) get an additional boost.
	`lrr.visualize(data, fb, ['ExternalRiskEstimate']);`
	 
	
	Credit inquiries
	The next two plots illustrate the dependence on the applicant's credit inquiries. The first plot shows a significant penalty for having less than one month since the most recent inquiry ('MSinceMostRecentInqexcl7days' = 0).
	`lrr.visualize(data, fb, ['MSinceMostRecentInqexcl7days']);`
	 
	
	The second shows that predicted risk increases with the number of inquiries in the last six months ('NumInqLast6M').
	`lrr.visualize(data, fb, ['NumInqLast6M']);`
	 
	
	信用查询次数的影

	Debt level
	The following four plots relate to the applicant's debt level. 'NetFractionRevolvingBurden' is the ratio of revolving debt (e.g. credit card) balance to credit limit, expressed as a percentage, and has a large negative impact on the probability of good credit. A small fraction of applicants (less than 1%) actually have NetFractionRevolvingBurden greater than 100%, i.e. more revolving debt than their credit limit. This might be investigated further by the data scientist.
	`lrr.visualize(data, fb, ['NetFractionRevolvingBurden']);`
	 
	
	The second 'NumBank2NatlTradesWHighUtilization' plot shows that the number of accounts ("trades") with high utilization (high balance relative to credit limit for each account) also has a large impact, with a drop as soon as one account has high utilization.
	`lrr.visualize(data, fb, ['NumBank2NatlTradesWHighUtilization']);`
	 
	
	债务水平的影

对于使用这个 AI 系统的银行贷款经理而言，他的关心重点是贷款决定的一致性，是否存在系统性的歧视。同样他也只希望一个指令获得解释，以便增强对模型的信心（即对模型普遍性的解释而不是单个案例的解释），而其依据是手边现有的案例为支撑（基于现有样本）对于特定结果的（事后）解释。依照上面的树形结构，我们自然能理解为啥银行贷款经理使用 ProtoDash 算法来寻求帮助。

有了这个决策，他按如下步骤开展：加载整理数据、运行算法、图示结论。

加载整理数据

```
heloc = HELOCDataset()
df = heloc.dataframe()
pd.set_option('display.max_rows', 500)
pd.set_option('display.max_columns', 24)
pd.set_option('display.width', 1000)
print("Size of HELOC dataset:", df.shape)
print("Number of \"Good\" applicants:", np.sum(df['RiskPerformance']=='Good'))
print("Number of \"Bad\" applicants:", np.sum(df['RiskPerformance']=='Bad'))
print("Sample Applicants:")
df.head(10).transpose()
```

```
Using Heloc dataset:  c:\users\ronnyluss\aix360\aix360\datasets\..\data\heloc_data\heloc_dataset.csv
Size of HELOC dataset: (10459, 24)
Number of "Good" applicants: 5000
Number of "Bad" applicants: 5459
Sample Applicants:
```

	0	1	2	3	4	5	6	7	8	9
ExternalRiskEstimate	55	61	67	66	81	59	54	68	59	61
MSinceOldestTradeOpen	144	58	66	169	333	137	88	148	324	79
MSinceMostRecentTradeOpen	4	15	5	1	27	11	7	7	2	4
AverageMInFile	84	41	24	73	132	78	37	65	138	36
NumSatisfactoryTrades	20	2	9	28	12	31	25	17	24	19
NumTrades60Ever2DerogPubRec	3	4	0	1	0	0	0	0	0	0
NumTrades90Ever2DerogPubRec	0	4	0	1	0	0	0	0	0	0
PercentTradesNeverDelq	83	100	100	93	100	91	92	83	85	95
MSinceMostRecentDelq	2	-7	-7	76	-7	1	9	31	5	5
MaxDelq2PublicRecLast12M	3	0	7	6	7	4	4	6	4	4
MaxDelqEver	5	8	8	6	8	6	6	6	6	6
NumTotalTrades	23	7	9	30	12	32	26	18	27	19
NumTradesOpeninLast12M	1	0	4	3	0	1	3	1	1	3
PercentInstallTrades	43	67	44	57	25	47	58	44	26	26
MSinceMostRecentInqexcl7days	0	0	0	0	0	0	0	0	0	0
NumInqLast6M	0	0	4	5	1	0	4	0	1	6
NumInqLast6Mexcl7days	0	0	4	4	1	0	4	0	1	6
NetFractionRevolvingBurden	33	0	53	72	51	62	89	28	68	31
NetFractionInstallBurden	-8	-8	66	83	89	93	76	48	-8	86
NumRevolvingTradesWBalance	8	0	4	6	3	12	7	2	7	5
NumInstallTradesWBalance	1	-8	2	4	1	4	7	2	1	3
NumBank2NatlTradesWHighUtilization	1	-8	1	3	0	3	2	2	3	1
PercentTradesWBalance	69	0	86	91	80	94	100	40	90	62
RiskPerformance	Bad	Bad	Bad	Bad	Bad	Bad	Good	Good	Bad	Bad

运行算法

1）预处理训练数据

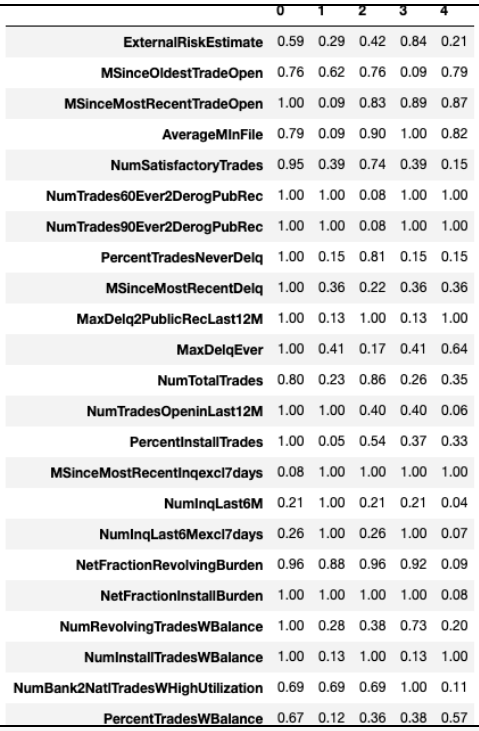
```
# Clean data and split dataset into train/test
(Data, x_train, x_test, y_train_b, y_test_b) = heloc.split()
Z = np.vstack((x_train, x_test))
Zmax = np.max(Z, axis=0)
Zmin = np.min(Z, axis=0)

#normalize an array of samples to range [-0.5, 0.5]
def normalize(V):
    VN = (V - Zmin)/(Zmax - Zmin)
    VN = VN - 0.5
    return(VN)

# rescale a sample to recover original values for normalized values.
def rescale(X):
    return(np.multiply ( X + 0.5, (Zmax - Zmin) ) + Zmin)
```

	N = normalize(Z) xn_train = N[0:x_train.shape[0], :] xn_test = N[x_train.shape[0]:, :]

2）定义和训练模型	# nn with no softmax
	def nn_small():
	    model = Sequential()
	    model.add(Dense(10, input_dim=23, kernel_initializer='normal', activation='relu'))
	    model.add(Dense(2, kernel_initializer='normal'))
	    return model
	# Set random seeds for repeatability
	np.random.seed(1)
	tf.set_random_seed(2)
	
	class_names = ['Bad', 'Good']
	
	# loss function
	def fn(correct, predicted):
	    return tf.nn.softmax_cross_entropy_with_logits(labels=correct, logits=predicted)
	
	# compile and print model summary
	nn = nn_small()
	nn.compile(loss=fn, optimizer='adam', metrics=['accuracy'])
	nn.summary()
	
	# train model or load a trained model
	TRAIN_MODEL = False
	
	if (TRAIN_MODEL):
	    nn.fit(xn_train, y_train_b, batch_size=128, epochs=500, verbose=1, shuffle=False)
	    nn.save_weights("heloc_nnsmall.h5")
	else:
	    nn.load_weights("heloc_nnsmall.h5")
	
	# evaluate model accuracy

	score = nn.evaluate(xn_train, y_train_b, verbose=0) #Compute training set accuracy
	#print('Train loss:', score[0])
	print('Train accuracy:', score[1])
	
	score = nn.evaluate(xn_test, y_test_b, verbose=0) #Compute test set accuracy
	#print('Test loss:', score[0])
	print('Test accuracy:', score[1])

3）验证现有数据集中类似案例将得到类似结论（即验证模型的一致性）	比如，我们选择案例#8，其结论是 Good，可以放款，其特征及如右图所示。 那么，数据集中有同样结论的其它案例有哪些？他们是否也有同样的特征分布？如果有，则表明这个模型具有一致性。 我们先获得所有结论是 Good 的案例；计算这些典型用户和#8 的相似度；我们获得了如下的相似度对比表，第 0 列是我们的选定案例#8，其它四个是数据集中有同样 Good 结论的其它案例。显然，过半数的特征的是接近的。仔细研究对照表，贷款经理发现，能获得 Good 结论的客户都是没有负债的客户，这个发现让贷款经理对系统的结论更有信息了。 ExternalRiskEstimate 82 MSinceOldestTradeOpen 280 MSinceMostRecentTradeOpen 13 AverageMInFile 102 NumSatisfactoryTrades 22 NumTrades60Ever2DerogPubRec 0 NumTrades90Ever2DerogPubRec 0 PercentTradesNeverDelq 91 MSinceMostRecentDelq 26 MaxDelq2PublicRecLast12M 6 MaxDelqEver 6 NumTotalTrades 23 NumTradesOpeninLast12M 0 PercentInstallTrades 9 MSinceMostRecentInqexcl7days 0 NumInqLast6M 0 NumInqLast6Mexcl7days 0 NetFractionRevolvingBurden 3 NetFractionInstallBurden 0 NumRevolvingTradesWBalance 4 NumInstallTradesWBalance 1 NumBank2NatlTradesWHighUtilization 1 PercentTradesWBalance 42

	
类似地，我们也可以对结论是 Bad（贷款申请被拒绝）的客户用同样的步骤做同样的研究	结果也同样是对于指定的样本客户，同样获得 Bad 结论的客户，其特征指标中过半数非常接近。 贷款经理仔细研究这些接近指标发现，这些被判 Bad 的客户，大都有轻微犯罪前科。为此，贷款经理在处理这类客户时就要额外小心。

	0	1	2	3	4
ExternalRiskEstimate	0.59	0.29	0.42	0.84	0.21
MSinceOldestTradeOpen	0.76	0.62	0.76	0.09	0.79
MSinceMostRecentTradeOpen	1.00	0.09	0.83	0.89	0.87
AverageMInFile	0.79	0.09	0.90	1.00	0.82
NumSatisfactoryTrades	0.95	0.39	0.74	0.39	0.15
NumTrades60Ever2DerogPubRec	1.00	1.00	0.08	1.00	1.00
NumTrades90Ever2DerogPubRec	1.00	1.00	0.08	1.00	1.00
PercentTradesNeverDelq	1.00	0.15	0.81	0.15	0.15
MSinceMostRecentDelq	1.00	0.36	0.22	0.36	0.36
MaxDelq2PublicRecLast12M	1.00	0.13	1.00	0.13	1.00
MaxDelqEver	1.00	0.41	0.17	0.41	0.64
NumTotalTrades	0.80	0.23	0.86	0.26	0.35
NumTradesOpeninLast12M	1.00	1.00	0.40	0.40	0.06
PercentInstallTrades	1.00	0.05	0.54	0.37	0.33
MSinceMostRecentInqexcl7days	0.08	1.00	1.00	1.00	1.00
NumInqLast6M	0.21	1.00	0.21	0.21	0.04
NumInqLast6Mexcl7days	0.26	1.00	0.26	1.00	0.07
NetFractionRevolvingBurden	0.96	0.88	0.96	0.92	0.09
NetFractionInstallBurden	1.00	1.00	1.00	1.00	0.08
NumRevolvingTradesWBalance	1.00	0.28	0.38	0.73	0.20
NumInstallTradesWBalance	1.00	0.13	1.00	0.13	1.00
NumBank2NatlTradesWHighUtilization	0.69	0.69	0.69	1.00	0.11
PercentTradesWBalance	0.67	0.12	0.36	0.38	0.57

对于受这个 AI 系统影响的客户而言（即贷款申请人），尤其是申请被拒的人，需要了解哪些因素是关键，从而他们可以采取行动改善自己的财经状况，以便日后有机会成功申请。其步骤也是类似：加载整理数据、运行算法、显示结论。这个就不赘述了。

要亲自上手试验，请参考 AIX360 网站 http://aix360.mybluemix.net。

576. 神经拟态计算为何被认为是下一代 AI？

——杜克大学陈怡然教授、浙江大学唐华锦教授、英特尔中国研究院宋继强院长热议神经拟态计算

神经拟态计算日益“火热”!

英特尔最近公布了在神经拟态计算的最新进展。英特尔 2017 年开发的异步脉冲神经网络 Loihi 芯片（SNN)，仅需单一样本便可学会识别 10 种有害气体的气味。以 Loihi 芯片为基础的神经拟态计算系统 PohoikiSprings 包含 1 亿个神经元，堪比小型哺乳动物的大脑容量。

神经拟态计算或类脑计算或神经形态计算，指的是机器模拟人脑神经机制和运行方式的有关计算。脉冲神经网络（SNNs）是神经拟态计算中一种全新的模型，可以模仿人脑中自然神经元网络的方式将计算模块重新分布。颠覆传统冯-诺伊曼的硬件加软件架构，实现人脑的智能功能，神经拟态计算被认为是引领下一代人工智能的主流计算模式。

三位学术界大咖在这次行业对话中着重对神经拟态计算的独特优势，与目前主流的深度学习的对比，如何进一步突破，以及未来应用方向进行交流。

解决现有 AI 的挑战。目前人工智能正面临诸多挑战，其中一项就是对于能源的大量需求，造成对生态的污染问题。

仅仅训练一个 AI 模型，消耗的电力（产生的碳排放量）相当于 5 台美式轿车整个生命周期的碳排放量。可以说目前的 AI 模式对生态环境构成了一定的威胁。作为下一代 AI 芯片，神经拟态计算能很好解决这一问题。

宋继强：神经拟态计算在算法以及芯片的设计上可实现以低一千倍以下的功耗去完成同样效果的模型训练。

唐华锦：芯片体积小、功耗低，符合生物进化最本质优势。

神经拟态计算优势：功耗低、神经元的智能性和自主性（最大优势）不是单纯解决一个数据训练、模式识别问题，而是解决多模态感知、非结构化信息的感知和推理。

陈怡然：神经拟态计算比目前技术更加安全。可以通过不同信号相互间连接做得更“鲁棒（Robust）”，这对于外在的攻击或不友好操作可更有效进行保护。

神经拟态计算与深度学习的关系。

神经拟态计算是否在不远将来取代深度学习？

Gartner 调查报告预测：到 2025 年神经拟态计算有望取代 GPU，成为下一代 AI 主流计算形态。

宋认为，神经拟态计算和深度学习的关系是兼容并蓄，不是取代。对深度学习已擅长模拟人类视觉或自然语言交互的任务，还是让深度学习的网络去模拟。在其他方面，如 Loihi 芯片做的嗅觉方面，还有机器人操控、多模态甚至跨模态之间的知识存储，可用神经拟态计算去实现。

唐认为，在一些特殊场景中，如并不需要太精确的计算结果，而需在一个实时环境中给出一个鲁棒响应时，神经拟态计算有绝对优势。

陈认为，目前两者实现的功能没有特别大的不同。神经拟态计算具有鲁棒性及实时性优势，这些只是在深度学习上提升而不是技术上突破。

英特尔现在成立神经拟态研究社区（INRC）期望在应用方面有所突破，抓应用场

景落地。

英特尔已研制出 PohoikiSprings（1 亿神经元）神经拟态计算机样机（非诺架构）并投入试用。与传统计算机比，运算速度提高 1000 倍，能耗降低 10000 倍。

神经拟态计算未来前景切入点:

①一个是非结构化数据，实时性要求高的场景，

②多模态的、实时的场景（如机器人、无人机），

③要持续学习、自适应学习的场景。

577. 清华沈向洋教授结合创新谈人工智能和开源

清华大学双聘教授沈向洋最近在 CNCC2020 圆桌会议上谈到人工智能和开源的问题。

他说：我们的创新第一是人工智能，最重要的事情就是要做可解释的人工智能。还有，我们为什么需要拥抱开源？

他结合自己过去十几年在微软工作时思考的几个大方向，目前有三个方向：第一个是人工智能，最重要的事情就是要做可解释的人工智能，如今深度学习发展很快，但可解释性这边进展较为缓慢。第二个是量子计算，这个路途还很遥远，微软从很早以前就走一条所谓的拓扑量子计算这样的路，但真正的量子计算器还有相当长的路要走。第三就是微软一直在推的混合现实。大家都很期待苹果之后发行的手机，到底 AR 是怎么样的，5G 上来之后，会有什么样的变化。这些我都非常期待。

我个人则希望做一些 AI 和神经科学之间的研究，神经科学研究实际上还处在早期阶段，数据不够，也做不了太多实验。我们能否解决一些真正的问题，例如阿尔兹海默症，中年忧郁症，儿童自闭症，这些都是大脑出现了问题。目前 AI 在学人脑，那么 AI 是否能反过来去帮助解决人脑的问题？也许有机会，清华的学生也好、其他人也好，大家能一起做一些事情。

我前面提到了，可解释性的人工智能有必要做。因为你如果不理解它，就很难去下判断，即这件事该不该做。

在谈到开源时，他说：创新就要做到极致，用开源的方式培养未来。他提出：为什么我们需要拥抱开源？

他说：过去 40 年，中国的科技发展非常大，因为出现了两个东西，一个是互联网，一个是开源。开源这件事情，对我们影响非常大。如果用开源的方式培养人才、提高水平，不要总提高 10 倍，只要提高 2 倍，对社会效率影响就会非常大。

578. 论文：审查对可解释的人工智能的需求

（Reviewing the Need for Explainable Artificial Intelligence）

作者：Julie Gerlings，Arisa Shollo，Ioanna Constantiou

单位：哥本哈根商学院

地址：https://arxiv.org/ftp/arxiv/papers/2012/2012.01007.pdf

发表时间：2020 年 12 月 2 日

人工智能应用程序在组织和社会中的普及推动了有关解释人工智能决策的研究。

可解释的人工智能（XAI）领域正在以多种方式提取信息并可视化人工智能技术的输出（例如深度神经网络）而迅速扩展。但是，我们对 XAI 研究如何满足可解释人工智能的需求了解有限。

本文作者对 XAI 文献进行了系统的回顾，重点关注与 XAI 有关的目的、定义和行为，并确定了关于 XAI 如何解决黑匣子问题的四个主题辩论。其次，作者从社会技术角度评估辩论，确定了两种未来的研究途径：a）利益相关者方法的必要性，并认识到不同利益相关者具有不同的可解释性需求；b）对可解释性需要整体看法和共同考虑社会问题以及 XAI 的技术、过程和结果方面，以及事实和讲故事方面。他们认为，要推进 XAI 的理论和实践，信息系统（IS）领域需要进行经验研究，以显示不同的 XAI 框架如何满足不同的利益相关者需求。

最后作者基于对 XAI 奖惩机制的这一批判性分析，作者将这些发现综合到未来的研究议程中，以进一步发展 XAI 知识体系。

其中 4 个主题介绍

在本节中，我们介绍了通过分析文章库而出现的四个主题辩论。

1） 激发对 xAI 的需求

探索有关 xAI 的最新文献和该技术的目的,我们观察到关于 xAI 定义的概念差异。基本概念有各种解释，例如解释与解释及其相关概念。 一些研究人员可以互换使用这两个术语，而另一些研究人员则描述了两个概念块之间的差异。

Miller 阐明了如何将社会科学中的解释视为两步过程，包括：a）认知过程，描

述事件原因的解释，其中选择了原因的子集作为解释（解释），以及 b）以交互方式在解释者和被解释者之间传递知识的社会过程。 而 Brandão 等的立场是将“好的解释”描述为一种解释，其中解释者理解了解释对提出要求的人的意义，因为他们强调有必要调查其对开发商和其他研究人员的意义的社会意义。

正如布赖恩（Brian）和科顿（Cotton）指出的那样，可解释性和可解释性的术语（及其变体）相互交织，而且在其定义中仍然很混乱。“解释与可解释性的概念密切相关：如果系统的操作可以被人类理解，则系统可以解释，通过内省或通过详尽的解释。”

其他学者，则采取更为务实的观点，认为“解释”更接近于模型的发展，并且与“黑匣子”模型相反，在黑匣子模型中，人们寻求对机制的直接理解。模型的工作原理是可解释的机器学习的目标。

其他人则将解释定义为向人类解释或以可理解的术语呈现的手段，并以人类如何解释信息的方式指导研究。廖等人考虑到不同的用户需求，主张采用更加多样化的 xAI 方法。 他们将 xAI 描述为“……举一个例子，解释 ML 分类器做出的预测的最流行的方法之一，因为许多 XAI 算法都在努力做到这一点，它列出了对模型的预测有最大权重的特征”对开发人员而言，这可能具有很高的价值，但对于普通的外行而言却不然。

这些不同的定义表明，在 xAI 领域需要进一步的概念对齐。 以下各节介绍了 xAI 系统的主要驱动程序。

1.1） 产生信任，透明和理解。

产生信任是 xAI 的主要推动力，并且与透明度密切相关。DARPA 的 XAI 计划促进了对 xAI 的需求，因为我们需要进一步了解，信任和管理新兴的人工智能机器。沿着这些思路，进行了大量的工作和研究，重点是从模型中提取信息或构建更简单的模型，以期实现透明，理解并从而建立对模型的信任。Gilpin 等认为：“...能够概括神经网络行为原因，获得用户信任或产生有关其决策原因的见解的模型...”。与 DARPA 一起，机器学习性能和使用的普遍增长促使人们寻求对模型的更好理解，以增加信任度，从而在业界增加机器学习的使用。此外，米勒认为，两种互补的方法将产生更加透明，可解释和可解释的系统，从而使我们更加有能力理解和信任模型：1）可解释性和可解释性被理解为人类对解释的理解程度在给定的上下文中；以及 2）对人（目标受众）的预测（决策）的解释。多数技术 xAI 方法旨在从模型（可能是神经网络或随机森林）中提取信息，例如特征重要性，相对重要性得分，敏感性分析，规则提取或其他方法以产生更大的透明度。这些 xAI 方法和框架主要是从透明性的感知出发，可以提高理解度，从而增加信任度–或相反，“黑匣子”模型不可信任。

很少有论文在所介绍的技术模型中包含社会技术方面的内容。然而，很少有人能解决利益相关者理解输出的障碍，其中包括将输出作为解释的社会技术方面的考虑，HCI 困境以及解决由开发者（庇护犯人）创建的为开发者提供解释的风险。例如，Zang 和 Zhu 提出了一种图形逻辑（或符号逻辑）来简化对卷积神经网络

（CNN）的理解，而不仅仅是信息提取。而穆勒等。可视化用于确定狼的沙哑以通过 LIME 测试参与者的像素。通过这种方式，他们通过测试参与者对他们是否信任该算法的解释来测试对人类理解的需求。此外，文献强调，生成解释的 xAI 框架是由开发人员或技术人员构建的，专注于提取数据中的计算问题，这不一定能解决信任问题。

但是，许多概念性论文呼吁进行跨学科研究，并讨论了需要更多关注人类理解或可解释性的问题，而不仅仅是透明度。

1.2） 确保合规，遵守法规和 GDPR 法律。

对新法规和 GDPR 法律的众多反应之一就是要求 xAI 不仅向用户提供解释，而且向整个社会提供解释。这以及其他法规，使得从业者和行业迫切需要加大投资以解释不透明的模型。GDPR 法规和“解释权”在研究和行业中引起了极大的轰动，将它们引向 xAI-作为合规性的可能解决方案。此外，一些研究者主张对 xAI 本身进行监管，或者为确保 xAI 的负责任使用而制定标准或质量措施的可能性，并避免建立有说服力的模型，而不是可以解释的模型。在 Gosiewska 和 Biecek 的实例中，很好地描述了构建有说服力的解释的谬误，其中示例是可加性模型如何导致对实例级别的解释产生误导性的指导，这一点得到了 Rudin 的支持，鲁丁反对最新的构建趋势（添加）可解释的事后“误导”解释。

1.3）为了履行社会责任，公平和规避风险。

特别是在医疗保健，临床和司法工作中，风险和责任是一个主要问题，因为它们

潜在地影响着人们的生活，而不仅仅是成本效益分析。 将责任分配给各个专业人士可以避免风险。因此，为专家（例如临床）推理开发心理模型，以更好地理解深度神经网络和不透明模型背后的推理。此外，最近在不透明模型中出现的歧视和累犯事件引发了关于确保模型性能的公平性以及对模型构建方式的更深入了解的辩论。在 xAI 文献中，招聘过程中的少数群体案例，COMPAS 系统中的累犯和普遍公平都在增加。

1.4） 建立负责、可靠和合理的模型进行论证

对 xAI 产生巨大吸引力的一个主题是，通过审查模型或创建其合法性的证据来确保模型的公平性和公正性。Adadi 和 Berrada[21]生成了一种可证明的方法，用以捍卫算法决策的公平性和道德性。除此之外，Abdul 等人提出了一种更新颖的生成 xAI 的方法，即建立基于因果关系概念的反事实解释。Liao 和 Anderson 在形式论证的基础上提出了生成基于论证的理由和解释的方法，这些方法为模型的更好推理提供了依据。最后，Ananny 和 Crawford 提出了一个关于透明度不足以管理和追究算法责任的讨论。他们声称，透明度不一定会建立信任，因为不同的利益相关者对系统的信任程度不同，这取决于他们对信息披露的时间和内容、以及信息的准确性和相关性的信任度。

1.5) 尽量减少模型性能和解释中的偏差和误解

模型的偏差和表现是 xAI 的一个重要驱动力，因为媒体经常报道模型的表现和性能不如人类，例如，在招聘过程中把合适的候选人筛选掉，或者未能识别出有色

人种。特别是在用训练集数据训练神经网络模型时，有偏差的训练数据会成为影响模型输出有效性的重大问题。除了有偏差的训练数据、变量选择和表征之外，我们自身的认知偏差还会阻碍我们对模型可视化输出的解释，因为我们往往会过度简化信息。

1.6) 能够验证模型并且验证 xAI 生成的解释。

针对有偏差的模型和模型不达标的表现，研究人员提出了四种深度神经网络（DNN）的评估方法，分别是:（1）与原始模型相比的完整性;（2）替代任务的完整性;（3）检测模型偏差的能力;（4）人类评估。其他研究人员则提出了一种用于评估可解释性的完全分类法，其中成本最高的是基于应用程序的方法，该方法需要对已实现的解释进行测试，并最终让用户对其进行测试。此外，他们还提出了一种以人为本的评估，例如在时间限制下哪种类型的解释是最好的。

2)完整性与可解释性困境

从评估 XAI 的争论中，出现了关于是否能够做出正确解释的争论。研究人员认为，对可解释性的需求源于不完全性会产生不同的偏见，并认为用户专业知识的性质将影响解释可能包含的复杂程度。许多其他研究人员主张在完整性和解释性之间的权衡取舍。我们应该谨慎对待这种折衷，因为人类对简单描述有强烈的特定偏见，这可能导致研究人员创建有说服力的系统而不是透明的系统。当健壮性较低时，人们会失去对解释的信任。

3)人的解释

关于如何解释和解释 AI 行为的许多研究是由构建 AI 的人而不是使用 AI 的人的需求所驱动的。用户可能具有的不同 AI 素养水平，而涉及利益相关者的多样性及其对 XAI 的不同需求的论文甚至更少。尽管有不同水平的 AI 素养和不同的学科领域，研究人员仍致力于开发以用户为中心的概念框架。只有少数几篇论文讨论了关于 XAI 生态系统的不同类型的角色和利益相关者，并认为一种解决方案可能不适合所有不同类型用户的目的，但我们需要包括利益相关者的背景，背景和知识，产生可以理解的解释。

4)技术产生了 XAI

近年来，在寻找打开臭名昭著的黑匣子的过程中，已经提出了许多不同的方法来构建更透明以及可解释的模型。可以分类如下:

本质透明: ML 模型具有更简单的特征，但不如其他更高级的模型（线性回归，逻辑回归，决策树）更精确

与模型无关的 XAI 框架: 这些通常具有事后特征，这意味着它们旨在适应任何模型类型，并依赖简化模型的技术，显示特征相关性估计，可视化模型或生成输出的本地替代模型。 这些框架的共同点是它们产生某种视觉输出，以便于理解。

579. 标题: 在微博上下文中产生可解释模型的歧视性表达

(Discriminatory Expressions to Produce Interpretable Models in Microblogging Context)

作者: Manuel Francisco, Juan Luis Castro

机构: 西班牙格拉纳达大学计算机科学与人工智能系（Department of Computer Science and Artificial Intelligence, University of Granada, Spain ）

链接: https://arxiv.org/pdf/2012.02104.pdf

发表时间: 2020.11.27

社交网站（SNS）是最重要的交流方式之一。特别是，由于微博站点的特殊性（及时性，简短文本等），它们被用作分析途径。有无数的研究以新颖的方式使用 SNS，但是机器学习（ML）主要集中在分类性能上，而不是可解释性和/或其他优势度量标准上。因此，最先进的模型是黑匣子，不应用于解决可能产生社会影响的问题。当问题需要透明时，有必要建立可解释的管道。可以说，管道中最具决定性的组成部分是分类器，但这并不是我们唯一需要考虑的事情。尽管分类器可能是可解释的，但生成的模型过于复杂以至于无法理解，因此人类无法理解实际的决策。本文的目的是提出一种功能选择机制（该流程的第一步），该机制可以通过使用较少但更有意义的功能来提高可理解性，同时在要求可解释性的微博环境中实现良好的性能。此外，本文提出了一种根据统计相关性和偏倚来评估特征的排名方法。为了评估模型的分类性能，泛化能力和实际可解释性，实验小组对五个不同的数据集进行了详尽的测试。实验的结果表明，就准确性，概括性和可理解性而言，本文的建议是更好的，并且到目前为止是最稳定的。

580，标题: 脉冲神经网络的时间替换反向传播算法

作者: 加利福尼亚大学圣塔芭芭拉分校 Yukun Yang

地址: https://arxiv.org/pdf/2011.09964.pdf

公开时间: 2020 年 11 月

加利福尼亚大学圣塔芭芭拉分校于 2020 年 11 月在 arxiv 平台上公开了一篇名为《Temporal Surrogate Back-propagation for Spiking Neural Networks》的论文。论文的简介如下: 脉冲神经网络由于能耗低、与人脑的工作原理相似这两个优点而被大量研究。反向传播算法是一种很强大的训练人工神经网络的方法，但是因为脉冲行为是不可分割的，反向传播算法不能直接应用于脉冲神经网络。之前的研究使用了梯度代理和随机性来近似脉冲神经网络时间、空间方向上的 BP 梯度，它们的缺点是省略了重置机制在脉冲神经网络每个步骤之间引入的时间依赖关系。这篇文章以更好的完善理论作为研究对象，对这个缺失的影响进行深入的研究。研究发现，添加重置机制的时间依赖性，会使新算法对数据集的学习速率的调整更可靠，但在 CIFAR-10 等大型任务上并没有太大的改进。从经验上讲，重置机制的时间依赖性的好处不值得使用额外的机制来实现，在大多数情况下可以忽略。

581. 标题: 在线学习时态神经网络架构

（A Temporal Neural Network Architecture for Online Learning）

作者: James E. Smith

地址: http://cn.arxiv.org/ftp/arxiv/papers/2011/2011.13844.pdf

发表时间: 2020 年 11 月 27 日

文章提出了一种利用脉冲时间的脉冲神经网络架构，该架构在学习过程中仅使用每个突触的局部信号和全局出现的聚类特点来调整突触权重。该论文的总体目标是该架构的直接硬件实现，因此所有的架构元素都很简单，而且精度很低。低精度使得学习时间非常快，在 MNIST 数据集上的仿真结果表明，和其他在线学习方法相比，该架构在相似的错误率的情况下学习时间至少快了一个数量级。

582. 人工智能四项核心技术:

①算力，中美比较美国占优势

②算法，中美占平势（过去美国占优势）

③大数据，中国占优势

④应用场景，中国占优势（加上中国人多优势更突出）。

583. 研发可解释的机器学习

——COPU 志愿者

打破机器学习的“黑盒子”研发可解释的人工智能，已经成为世界（尤其是美欧）当前的一大亮点。

所谓机器学习或深度学习技术一般是不可解释的，或用以作业是不透明的。不可解释的机器学习也是一种人工智能（初级阶段的人工智能），在用以作业（以提高智能）或对之训练（以提高算力）时，由于其带有“黑盒子”、盲操作的缺陷，将致使作业或训练成果较差（或较弱），这时的不可解释机器学习也称为弱人工

智能（这是早期人工智能）。只有实现了可解释的机器学习（或给予机器学习模型以解释的能力），才有可能达致强人工智能。

为了打破机器学习中的“黑盒子”，导致使用机器学习作业及其训练透明化，变其不可解释为可解释，需要针对不同应用场景研发出各种不同、适用的可解释工具（包）或算法。使这些不同任务的解决方案或训练成果分别达到不同的目标:

①提高能效或绩效，提高性能或质量，

②提高判断、决策能力（或减低投资风险），

③开辟公平、信任或合规、遵法以及明责、打假等的新场

584. 答李院士问六：自然语言问题———王迪兴

问六：自然语言是人类思维话动的载体，如果自然语言是第一语言，数学语言是第二语言，计算机语言是第三语言，后一个比前一个更严格、更狭义。数学自身难以完全自治，数学的形式化借助于自然语言，计算机语言的形式化要借助于数学语言。因此，人工智能怎么可以反过来要用数学语言或计算机语言去形式化人类的自然语言呢？

答：自然语言、数学语言、计算机语言必须基于某种同构性才能建立，互相翻译、代偿、转换。没有同构性任何语言都无法相互沟通与交流，更不会产生有效认知结果。各种语言都是准完备的，都要符合逻辑一致性，基于逻辑一致性进行内涵和外延拓展，才会体现语言的有效性。任何广义语言都不能单一发挥作用，一定要有各种表达方式互补，如语言必须与文字、图形甚至手势互补发挥作用。语言

不存在谁形式化谁的问题，都是基于同构互补发挥作用。

计算机语言是基于自然语言派生的，但基于机器语言不能派生自然语言，因为计算机本身不具创造语言的功能，其语言不具主体地位，也没有自主社会化交流的需求，不存在计算机语言形式化自然语言的问题！

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一批可喜的进展，也面临着诸多的挑战。为了应对这些挑战，学术界涌现了丰富多彩的创新思路。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，根据笔者的分析，人工智能面临的最深刻最本质的挑战，是学科和时代的大转变所带来的大阵痛：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，便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？

智能与智慧是相联系又相区别的概念。**人类智慧**是人类为了生存发展的目的而运用知识去探索未来发现问题（**隐智慧**）和在此基础上变革现实解决问题（**显智慧**）的能力。隐智慧严格依赖于人类目的和思辨能力，只有人类能享有。而显智慧依赖于人类的智能操作，故称**人类智能**，可用机器来模拟。

人工智能是以人类智能(变革现实解决问题的显智慧)为原型、研究具有 智能水平的机器系统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

人工智能系统的具体形态多种多样，分别源于三大学派。(1) 1943 年发端 的以模拟人脑结构为导向的人工神经网络学派，(2) 1956 年兴起的以模拟人脑功能为标志的专家系统学派，(3) 1990 年前后问世的以模拟智能系统行为为特 色的感知动作系统学派。

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。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。这些问题的严重性表现在：

（1）智能水平低下，就**不够资格成为真正的人工智能**。

（2）没有整体人工智能理论，就表明**人工智能的研究远远没有上轨**。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，也远远不如整体成果。但是如何才能解决整体理论的建构？至今仍然众说纷纭。

3、人工智能存在问题的总根源：范式张冠李戴

人工智能的研究之所以会存在上述这些严重的问题，**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（不可恢复） 生命线和灵魂的各个子系统，就不再可能恢复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的“单纯形式化” 方法论。智能的决策能力根植于对研究对象的形式、内容、价值的全面理解，而**施行“单纯形式化”的结果，丢失了内容和价值这样的智能“内核”**，单凭 对表面形式的了解很难做出智能的决策。这是“智能水平低下”的根本原因。

“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，功不可没。然而，把它们用到人工智能研究领域，就用错了场合，变成了“张冠李戴”！

众所周知，方法论是为科学观服务的，有什么样的科学观就要求有什么样的 方法论。传统学科的方法论在人工智能研究领域产生了负面的效果，说明传统学科的科学观也不适合于人工智能的研究。

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两者一起，就决定了学科研究应当遵循的规范方式，称为学科的研究范式，简称范式**。

思想指导行动，任何学科的研究都必须遵循本学科的研究范式。以上的分析 表明：**人工智能的研究范式张冠李戴**（用传统物质学科的范式来规约作为开放复杂信息系统的人工智能研究）是人工智能存在各种问题的总根源。

4、人工智能研究范式的张冠李戴，是历史的必然

学科的研究活动（社会存在）都由它的范式（社会意识）所支配。但是，由于“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，直到如今都还**没有能够形成信息学科的范式，因此，信息学科的研究便沿用了业已存在的传统物质学科范式**。于是就造成了整个信息学科（含人工智能）范式的张冠李戴。

可见，人工智能范式的张冠李戴，并不是笔者的主观臆断，而**是由物质学科主导的学科体系向信息学科主导的学科体系（由工业时代向信息时代）历史性大转变所带来的大阵痛**，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桎梏，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的第三栏），但实际上却是遵循了传统物质学科的研究范式（请比较表1的第1栏和第2栏）。这就是“人工 智能范式张冠李戴”的具体表现。造成张冠李戴的深刻原因已在第4节阐明。

可见,只有颠覆和摒弃传统学科范式对人工智能研究的钳制,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。

这不是普通意义上的创新，而**是人工智能基础理论的彻底革命：**在学科 研究的最高引领层次上实施的人工智能研究范式颠覆性变革，是实现信息学科范式从无到有的历史性突破和确立，是人工智能理论的彻底转轨。

二、范式革命怎样创生通用智能理论？

1、发现“学科创生” 的一般规律：范式管控学科创生的全局全程

摆脱了传统物质学科范式对人工智能研究的束缚，确立了信息学科范式对人工智能研究的引领，全新的智能理论 — 通用智能理论 — 便终于破土而出。

创建“通用智能理论”的历程可以用笔者总结的“学科创生规律”来描述， 详见表2。

表 2学科创生规律

生长阶段	模块名称	模块要素	要素解释
探索阶段自下而上的探索	**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）
建构阶段自上而下的建构	**学科范式**（宏观定义）	科 学 观	宏观上明确“学科是什么”
		方 法 论	宏观上明确“应该怎么做”
	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”的整体研究方法
	学科规格（精准定格）	学术结构	基于“学科定位”的学科宽度规格
		数理基础	基于“学科定位”的学科深度规格
	学科理论（完整理论）	基本概念	基于“学科基础”的学科基本知识点
		基本原理	基本概念之间的相互联系

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，然后是自上而下贯彻范式的建构阶段。

探索阶段 特点是：各种学术背景的研究者们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的范式（科学观和方法论）。由于是经验性的摸索试探，常常发生“盲人摸象”现象，因此这一阶段经历的时间可能会很长：信息学科范式的形成便超越了半个多世纪。一旦达成了关于“范式”的共识，就可以自动转到建构的阶段。

建构阶段 特点是：根据探索阶段所总结的学科范式(即学科的宏观定义，包括学科的科学观和方法论)，就可构筑学科的框架（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）、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）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贯彻学科的范式。

以下各节的内容，将分别介绍笔者团队依照表2的“学科创生规律”在信息 学科范式引领下构建“通用智能理论”的系统成果。

2、研究和提炼“信息学科范式”

从当前的实际情况判断，无论是国内还是国外，人工智能学科应当遵循的范 式都还处在“探索阶段”。因为，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，共识罕有。

值得庆幸的是，由于特殊的学术兴趣、背景和经历，本文笔者早在半个世纪 之前就开始关密切注和潜心探索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大学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（信息的形式），丢了信息的内容和价值这两个核心要素，造成了“信息的空心化”。于是开始探索“形式、内容、价值”三位一体的“全信息理论”。26 年之后，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 年在研究人工智能的时候，又发现它被分解为 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，都受到传统学科方法论“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1996年的第二版、2002年的第三版、2005年 的第四版、2013年的第五版每次再版都深化了信息学科方法论的探讨。2014年， 笔者出版了另一部学术专著《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。最终，总结提炼成为本文表 1第三栏信息学科范式的标准表述，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宝贵成果。

3、范式革命：颠覆旧范式，确立新范式

颠覆传统学科范式对人工智能学科研究的统领地位，具体表现为：

（1）在信息学科的研究领域摒弃传统学科的科学观，后者认为：

（1-1）把研究对象看作是纯粹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）在信息学科研究领域摒弃传统学科的方法论，后者包括：

（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，具体表现为：

（1）在信息学科的研究领域确立信息学科的科学观，后者认为：

（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）在信息学科的研究领域确立信息学科的方法论，后者包括：

（2-1）信息生态演化方法

（2-2）形式、内容、价值三位一体的整体化方法 可见，两种范式相反相成。要让人们放弃传统学科范式，其实远非易事！

4、遵循信息学科范式，构筑全新“人工智能全局模型”

信息学科范式的科学观业已指明:人工智能是代表主体(人类或其他生物）的意志，在主体的驾驭下主体与环境客体相互作用而形成的不确定性信息过程。

其中，所谓“代表主体（特别是人类主体）的意志”，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”；在这个框架下实现主体所确定的目标。

根据这些要求，可以构筑图1所示的人工智能学科研究的全局模型：

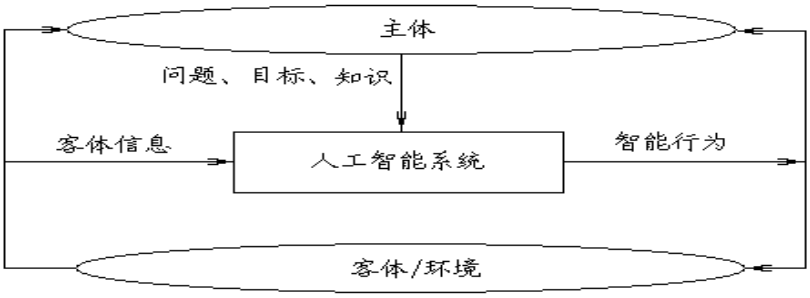


图 1根据信息学科范式构筑的“人工智能全局模型”

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、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，确实完全是自 然主体的目的，而不是人工智能“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**，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合作伙伴。

5、遵循信息学科范式，揭示智能生成机制，开创机制主义研究路径

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 允许被分割）方法来处置、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。

把信息学科范式方法论与图 1 的全局研究模型结合起来，就明确了人工智能系统生成智能的机制（注：“机制”与“机理”两者是同义语）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。

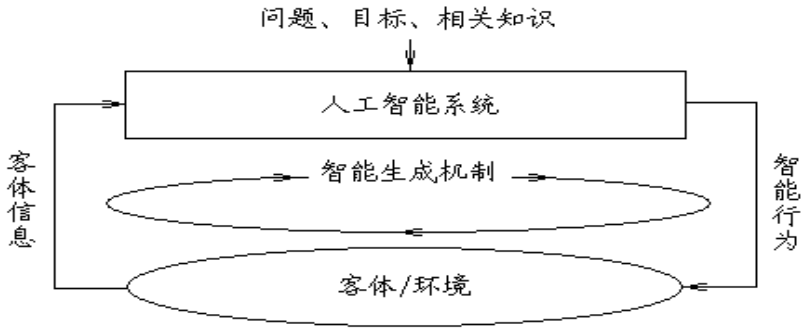


图 2人工智能系统的智能生成机制

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 的“客体信息”，它的结果是主体操作下由客体信息转换生成的“智能行为”，而它所遵守的约束条件则是由主体所规定的“问题、目标、知识”。

质言之，智能生成机制的实质是“**信息转换与智能创生**”，它的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**，恰如图3所示。

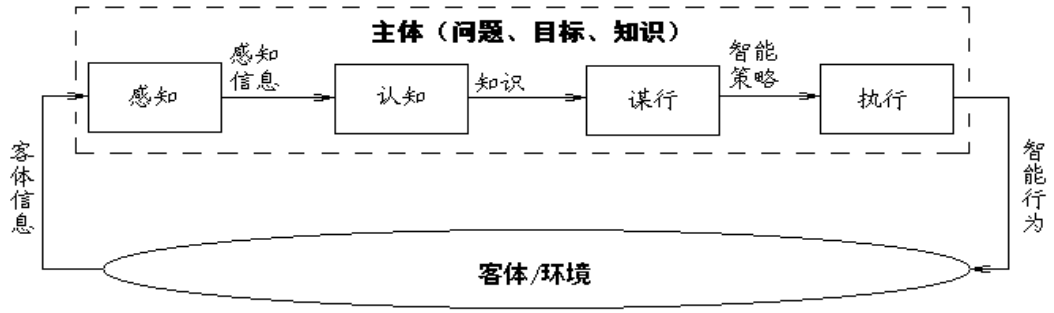


图 3智能生成机制“信息转换与智能创生”的具体化

可以证明，“信息转换与智能生成”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**，并由此把它的本质称为“**信息转换与智能创生定律**”。

普适性智能生成机制（机理）是人工智能研究的核心问题，也是人工智能研 究的根本路径。于是，我们把以“普适性智能生成机制”为基础的研究路径，称为“机制主义”研究路径。这是与现有人工智能研究路径截然不同的研究路径。

6、遵循信息学科范式，重审人工智能的学术结构

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，我们**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，竟把整个人工智能看作是计算机学科的一个分支。

7、遵循信息学科范式，重塑人工智能的数理基础

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，团队的何华灿教授创建了具有可调参数的“柔性（泛）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；与此同时团队的汪培庄教授创建了以因素为基元的“因素空间数学理论”，从而把原来互相离散的普通集合论、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数学理论。

8、遵循信息学科范式，重构人工智能的基础概念

在传统物质学科范式引领下，人们建立了一批人工智能的基础概念。但是， 由于接受了“纯粹形式化”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”并不是要求“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，智能理论最基础的概念 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。全信息的概念不是随 便说说而已，而是具有严格的生成机制，见图4：

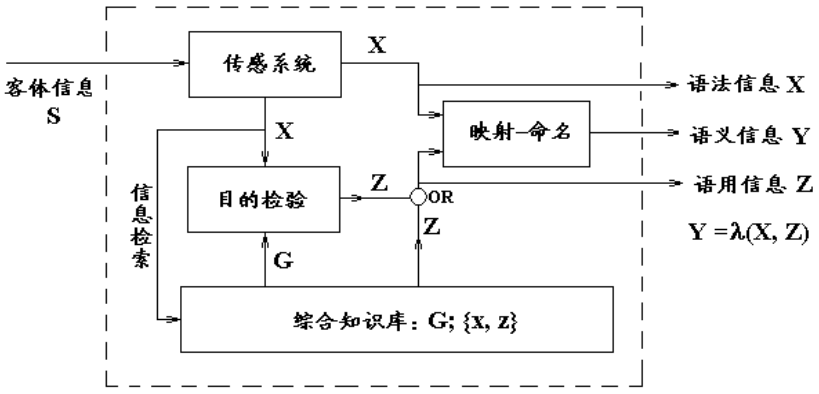


图 4全信息的生成机制

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“λ”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，也就理解了信息。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，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“全知识库”也比目前流行的“知识图谱”更为优越。

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、遵循信息学科范式，深挖人工智能的基本原理

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）“感知信息与知识→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。这既是人工智能的基本原理，也是人类智能的基本原理。

值得指出，“信息转换与智能创生定律”的深远意义还在于，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。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不能逾越的界限，信息转换与智能创生定律则告诉人们， 可以通过信息转换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。

10、遵循信息学科范式，创建通用智能理论

综合集成以上各项成果，特别是其中以“信息转换与智能创生定律”为旗帜 的机制主义研究路径（具体包含上述五项基本原理），可以创立“通用智能理论”，它的模型如图5所示。

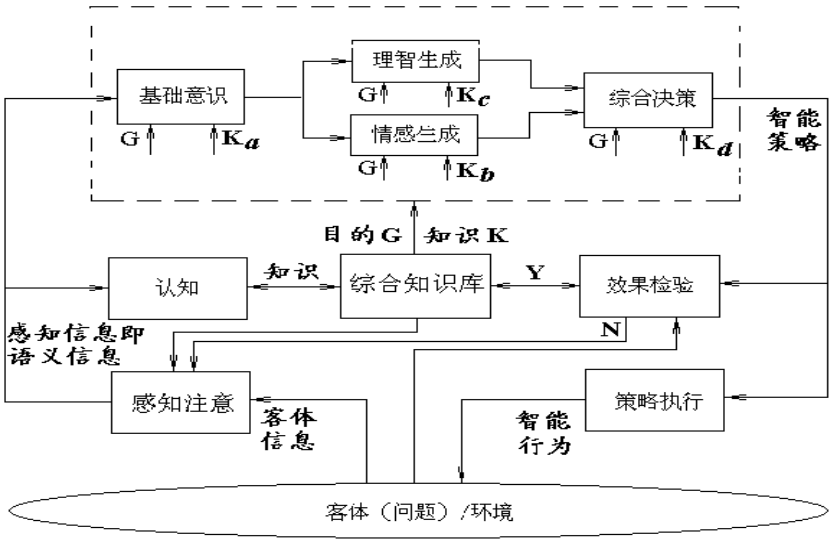


图 5通用智能理论模型

可以看出，图5的通用智能模型不但全面体现了信息学科范式（科学观和

方法论）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“认识世界和改造世界，并在改造客观世界的过程中也改造自己”的主客互动理论。

三，附带的话

人工智能的范式革命与通用智能理论的创生，涉及到许多深刻的科学问题和 哲学问题。为了节约篇幅，以下这些问题只列出题目而不做解释：

1、通用智能理论模型（图5）的实现原理已经落地。

2、通用智能理论可以克服现行人工智能存在的各种问题。

3、通用智能理论系统可以孵化出各种应用的人工智能系统。

4、把孵化出的各种人工智能系统用于各行业，我们规划为“洛神工程”。

5、目前正在寻求合作，把理论转化为通用智能孵化平台和现实生产力。

阅读笔记:

敬请关注联盟微信公众号
COPU开源联盟

中国开源软件推进联盟秘书处　　电话：+86 010-88558999
联盟公共邮箱：office@copu.org.cn　　联盟官网：http://www.copu.org.cn
地址：北京市海淀区紫竹院路66号赛迪大厦18层