评人工智能如何走向新阶段

陆首群

2021.4.30

围绕《评人工智能如何走问新阶段》这个主题，国内外专家以跟帖留言方式展开大讨论，从2019年8月8日始，迄今（2021年4月30日）历经1年8个多月，已收到跟帖704条。这些跟帖显示了70年来人工智能发展的轨迹，概括了近年来全球人工智能发展的最新成果，正在探索人工智能的未来走向。

早年兴起的人工智能：机器学习／深度学习／增强学习的算法是一种强大的数据分析工具，但机器学习／深度学习也是有缺陷的，深度学习本质上是一项暗箱技术或盲模型，其训练过程不可解释、不可理解、不可控。中外都有人认为，机器学习／深度学习算法的潜力已近天花板，人工智能又步入了一个低潮循环。对此业界有争议，持不同看法的人认为，说机器学习／深度学习潜力耗尽言之过早，说人工智能步入低潮也不符合现实！举看作当今世界研发人工智能重镇之一的谷歌为例，在其发表的谷歌在《2020年人工智能10大领域的发展与成就》中可以看出：机器学习算法仍是其研发的重点，占70%，它特别强调：机器学习／深度学习算法还是一个非常话跃的研究领域，对机器学习／深度学习的算法与模型要加深理解、改进、创新，将继续系统地重塑其算法。

今天世界人工智能的繁荣场景都是基于机器学习／深度学习的，如医学诊断、语音识别、图像识别、自然语言处理、自动驾驶、全球首架六代机、全新一代抗生素、大量新材料、新药物等的研制。

最近沈向洋教授在深圳《人工智能与机器人国际研讨会》上作“从深度学习到深度智能”的报告，他指出当前的机器学习／深度学习属于狭义的人工智能或

弱人工智能，需要综合各种来源的知识，期望其能够对世界上正在发生的事情进行推理，能够像人类一样，在一种语境中学习在另一种语境中应用，发展强人工智能。对于如何从机器学习／深度学习走向强人工智能？沈教授说他正在思考，报告中提出了三个方面走向：

一是构建大规模的强机器学习仿真器（不仅用于游戏还用于自动驾驶等复杂系统）

关于这条建议我先不予置评。

二是对机器学习本质的深度理解，从优化功能开始，思考我们从里面真正学到什么？

这一条与我在前面谈到的谷歌的思路是一致的。

三是基于（深度）神经与符号的混合模型

这一条与我们在跟帖中介绍的一批 AI 资深专家对发展新一代通用人工智能（或强人工智能）的设想是相似的。

接下来，我们也来谈谈一个方面的走向：

如何打破机器学习黑盒子，研发可解释、可推理、可控的人工智能（这在我们发布的人工智能国内外跟帖中已提到有十几例这方面的理论和实践），未知沈教授如何看待这个走向？

谈到人工智能的出路在何处？从已发表的 704 条跟帖中可以列出 4 条路径：

1）打破机器学习的黑盒子，研发可解释的人工智能

这条研发的路径已成为今天全球研发的热点。

2）基于异步脉冲神经网络的神经拟态（类脑）计算系统

英特尔、浙江大学研发了基于类脑芯片的神经拟态计算系统（非冯-诺伊曼

计算架构）

值得注意的有关专家指出，研发基于脉冲神经网络（类脑）芯片之上的神经拟态计算系统主要用符合自然原理的硬件来实现，软件代替不了硬件，但迄今用硬件来实现的成果较少，这也是类脑人工智能实现突破性进展有限的原因

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

这项研发还差最后一公里。在大规模语义网络中还存在一些短板，如未能解决常识、专家经验、语义理解等问题

4）脑机接口的理论和实践

目前这项在国内外已有几十例试点，主要用于癫痫、中风病人辅助治疗。

在讨论人工智能未来走向何方？专家们在704条跟帖中引发了学术争论。一些人工智能资深专家认为，原来人工智能的发展路径离不开三大学派：①基于数据统计的连接系统的结构主义学派，②基于逻辑推理的符号系统的功能主义学派，③基于模拟智能生物行为的感知动作的行为主义学派。而迄今人工智能发展中，三大学派各自取得了（或将继续取得）不少精彩成果，但三派争霸无法统一，均存在片面性，难以通往新一代通用人工智能（强人工智能），他们分别提出了通用人工智能的新理论（在704条跟帖中可看到）。

国内外人工智能跟帖留言 646-704 条:

646. IBM 向 LF AI 捐赠 AIX360 项目，助力可解释 AI 实践

人工智能的普遍应用需要达成知其然，也知其所以然。这是可解释 AI 的使命。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，需要通过“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。

所谓“事后分析解释”既可以用来理解所使用数据，也可以用来理解使用特定数据所训练出的模型。对于前者,可以采用 DIP-VAE 算法以提取哪些是最有效特征，也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，从而帮助用户理解系统决策的整体合理性。局部解释就特定案例进行分析，找出影响该模型做出该项结论的关键要素。

AIX360 是对于这些训练数据和模型建立事后分析解释的工具包。IBM 的研究人员在 http://aix360.mybluemix.net 上给出了一个银行信贷决策系统的可解释实践。该 AI 系统基于美国的公开、真实的个人金融数据集(FICO HELOC Dataset)来辅助决策是否批准某项贷款申请。

本项目的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，从而建立如下的规则：

“拥有少于 5 个户头或者拥有多过 5 个户头且每个斧头负债超过 1000 美元的客

户是高风险客户”银行的主管当局依据自己多年的工作经验，十分认同这样的规则，从而认可本系统所能给出的贷款决策建议。

银行信贷经理关心的是这个系统给出的决策建议是否有系统性歧视（比如种族、肤色、年龄、性别、价值观等)。为此，他要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。经过演算分析，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。这给了他很大的信心。通过分析这些共同特性，他发现未能获得贷款的申请者大多有轻微违法犯罪记录。这可能有过于严苛并有失公允，要额外小心处理。

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

647，可解释 AI 的算法选择和执行步骤

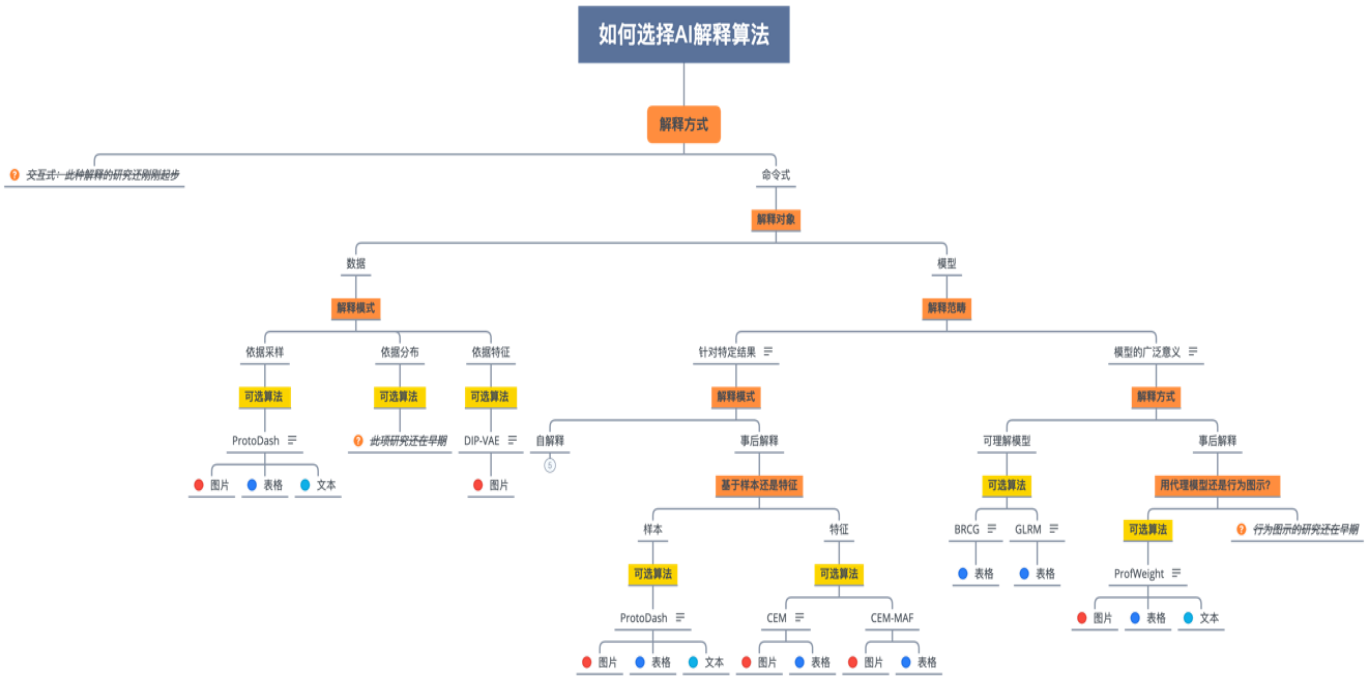
当前，采用机器学习、深度学习进行 AI 决策的系统往往有黑盒子特点。可解释 AI 通过“事后分析解释”(post-hoc interpretation)来帮助用户打破黑盒子、建立对于该系统决策的合理信心。不同的人在不同的场合、不同的背景下，对于可解释 AI 有着不同的需求。

AI 系统的开发者，其目的多半是如何提高系统效率。AI 系统的使用者，则是需要建立对于这个系统所做决策的信心，能够放心、安心采纳其建议。而市场或者业务的主管乃至监管当局主要关心的是如果确保系统性的公平。最终受影响的用户则需要能够理解影响结论的主、次因素，从而未来能够有所作为。

下面的决策树概括了当前可解释 AI 的实践现状。

目前的可解释AI可以胜任事后分析解释,还不能完成交互型探究式的分步解释。

不同的数据形式（图片、表格、文本）、不同的解释范畴（全局还是局部、特征分析还是案例分析）需要不同的解释算法。



要解释一个 AI 系统，从算法角度，大致分为三个步骤：

- 获取并加载、整理数据
- 依据需求选择、运行算法
- 整理并显示可理解结论

例如，一个银行贷款决策系统开发团队的项目经理要向甲方说明系统的功效，即无需经年累月的培训和项目经验积累，本系统能帮助贷款部的任何工作人员依照所积累的数据及时做出和资深经理依据多年经验同样的贷款决定。

项目经理选择了采用 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。为此，他按如下步骤展开工作：

获取和加载、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

选择和运行算法	
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

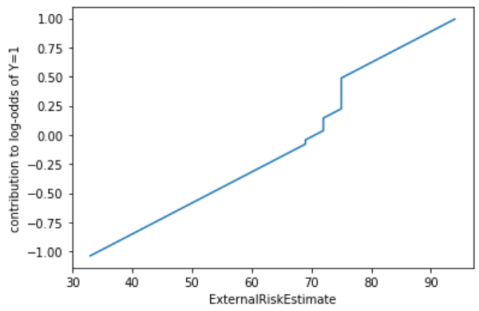
（图形）显示结论，如以	外部风险预估

GAM 图示 LogRR 的结论

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']);
```

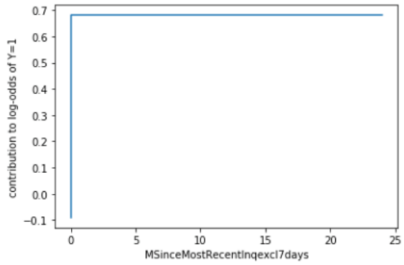


信用查询次数的影响

Credit inquiries

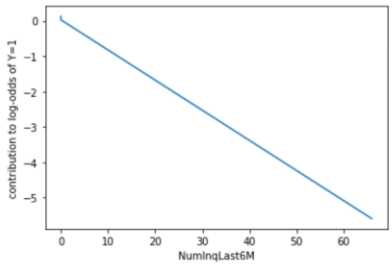
The next two plots illustrate the dependence on the applicant's credit inquiries. The first plot shows a significant penalty for having less than one month since the most recent inquiry ('MSinceMostRecentInqexcl7days' = 0).

```
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']);
```

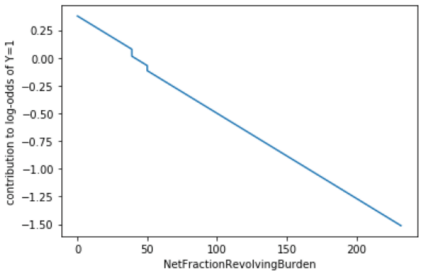


债务水平的影响

Debt level

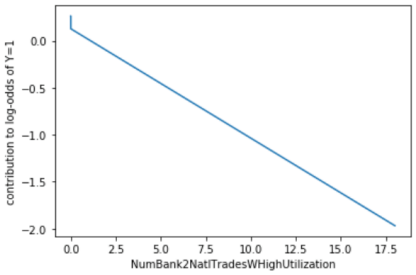
The following four plots relate to the applicant's debt level. 'NetFractionRevolvingBurden' is the ratio of revolving debt (e.g. credit card) balance to credit limit, expressed as a percentage, and has a large negative impact on the probability of good credit. A small fraction of applicants (less than 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']);
```



以上图表的结论和资深经理的经验不谋而合，项目团队获得了甲方的认可。

相关算法的详情和代码、如何获取 FICO HELOC 数据集、亲自上手试验这些算法，请参考 AIX360 网站 http://aix360.mybluemix.net。

648，可解释人工智能（XAI）教程系列

AAAI 发布，2021.2.3，可参阅：

https：//xaitutorial2021.github.io/#
http://www-sop.inria.fr/members/Freddy.Lecae/presentation/aaai-2021/xai-turial.pdf

可解释人工智能只在通过象征性人工智能与传统机器学习优势结合来应对此类挑战。多年以来，各种不同的机器学习社区都以不同的定义、评价指标和动机、结果等表现该主题。本教程主要是机器学习和 XAI 相关方法的介绍（分成五大部

分）：

1）从理论和应用的角度来描述和激发对可解释人工智能的需求（简介）

2）可解释人工智能（不仅仅是机器学习）

3）知识图在可解释人工智能中的应用

4）可解释人工智能应用程序和课程

5）可解释人工智能工具，编码实践和研究挑战

649，严防中国人工智能赶超美国

《网易新闻》2021.2.25

当地时间 2021 年 2 月 23 日，在美国国会参院一场听证会上，谷歌前董事长埃里克-施密特发言：“在人工智能领域，美国只领先中国一两年，而非五年十年”，“中国在人脸识别方面遥遥领先”。

施密特此前在多个场合说“要不惜一切代价在高科技方面击饭中国”，并提醒美国政府“紧盯中国”。

650，新型粒子加速器光源“稳态微聚束”验证试验成功，有望解决自主研发 EUV 光刻机中最核心卡脖子难题

清华大学工程物理系教授唐传祥研究组与来自亥姆兹柏林材料与能源研究中心（HZB）及德国联邦物理技术研究院（PTB）的合作团队，在《Nature》上发表题为《稳态微聚束原理的实验演示》（Experiment ul demonstration of the mechanisms of steady-stalemicrobunching）的研究论文，报告了一种新型粒子加速器光源“稳态微聚束”的首个原理验证试验，基于 SSMB 的 EUV 光源有望

解决自主型研发 EUV 光刻机中最核心的卡脖子难题。

651，陆汝钤院士支持钟义信先生的《通用人工智能基础理论研究》课题

陆汝钤（2020-07-30）

陆汝钤，中科院院士，是吴文俊人工智能最高成就奖获得者

（全文如下）

钟义信先生建议的《通用人工智能基础理论研究》课题是一个非常有意义的研究课题，我表示完全支持。钟先生在这方面已经有过深入的研究，其中和其他几位国内人工智能界前辈共同开发的机制主义人工智能理论已经广为人知。该理论的宏伟目标是把传统的几派人工智能主流理论统一起来。另外，不久前钟先生还提出了智能发生机理的重大理论问题并组织了座谈会。这一切都是对人工智能基础理论的重要贡献，并且为申请和执行“通用人工智能基础理论研究”课题奠定了基础。

我觉得这个项目最大的意义在于在人工智能领域举起了“基础理论研究”的大旗。为了说话简便，下面就用 AI 来代表人工智能。现在 AI 的热度可能是历史上最高的。大力发展 AI 已经成为我们的国策。国务院发了文件，提出了指导性的 AI 发展规划。科技部拨出专款，资助重大 AI 项目的研究。大大小小的 AI 企业如雨后春笋般出现，各种各样的 AI 产品纷纷问世。而 AI 基础理论研究的声音却相对显得薄弱。尽管国务院文件也提到了 AI 基础理论研究，但是这类研究的现实情况却不容乐观。以钟先生为首的专家团队提出《通用人工智能基础理论研究》课题旗帜鲜明地打破了这种冷清的局面，这是我们全力支持本课题的根本原因。

从研究内容来看,钟先生团队提出的几个AI基础理论问题都是很有意义的，值得深入研究。同时我们也认为，本项目的研究可以不限于申请书里提到的那些问题。事实上，AI研究面临的基础理论问题是相当广泛的。比如，大家都在热衷于不断发明新的AI产品和AI技术，并且不断地宣布：AI又能够做什么新的事情了。但是好像没有人研究过AI不能做什么。这不禁引出一个问题：AI是万能的吗？AI的能力有极限吗？是不是可以提问：存在不存在不可AI的问题呢？这里我们把‘不可AI’的问题定义为这样的问题类：对这一类问题，不存在而且永远不会有比人工更好的AI解决办法。或者减轻一点说：有无这样的问题类，对它不存在而且永远不会有不需要人工参与的纯粹AI解决办法？

如果说不可AI问题是AI问题求解的上限，则也可以探寻AI问题求解的下限。在AI发展的历史上许多曾经被认为是AI的发明现在可能不再被列入AI的范围了。比如说某些计算问题和某些逻辑推理问题。这些问题可以称之为“不算AI”问题。那么一般地说，什么问题不算AI问题呢？谁能画一条界线？我想一个问题算不算AI问题，每个人心里都有一杆秤。如果把它作为一般性的问题拿出来讨论，很可能是会有较大分歧的。

这样的问题还很多。例如智能的本质是什么？AI和人类智能有什么本质的区别？混合智能又是怎么回事？所谓强AI的真正意义是是什么？从理论上看能达到吗？等等。总之，许多问题有待回答，使我们对这个项目寄予厚望。国内外的AI前辈们对AI本质提出了许多学说。这些学说各有各的道理，都在历史上发挥过重要作用。希望它们在这项研究中能得到融合和升华，催生出更具说服力的新理论。从长远来说，任何先进的理论最终都会促进社会福利和人类进步。理论的基础性越强，它的效应就越不会来得太快。所以我个人建议在目前的项目建议中

不必过分强调本项目在工程技术和国计民生上的作用。正是因为社会上有一种对AI 基础理论研究重视不够的现象，我们才更要名正言顺，理直气壮地强调本项目就是要做 AI 基础理论研究。以上是我们的想法和建议，仅供钟先生和项目团队参考。

652，《金融时报》网站 3 月 2 日透露的信息，美国国会授权成立的人工智能安全委员会警告说，美国有可能失去芯片优势，在人工智能领域中国是一个强大的竞争对手。该委员会联席主席是：谷歌前 CEO 埃里克-施密特，美国国防部副部长鲍勃-沃克。

埃里克-施密特说，我们对台湾芯片的依赖非常接近失去驱动我们的公司和军队的微电子尖端优势，鲍勃-沃克说，如果中国大陆收复台湾，那对我们将从领先两代变成落后两代。该报告还说，尽管我们私营部门和大学在人工智能领域处于领先地位，但美国仍未就即将到来的时代作好准备，在人工智能领域，中国拥有挑战美国技术领先地位、军事优势及在全球更广泛地位的能力、人才和野心。

653，陆总：这是我们研究工作的一个简介，请批评指正！

钟义信，北京邮电大学

2021.03

“机制主义人工智能基础理论”简介

《机制主义人工智能理论》及其逻辑与数学基础（合称为机制主义人工智能基础理论）是实现了“人工智能基础理论重大突破”（国务院规划的第二步战略目标）**的重要理论成果**。

突破的关键是抓住了学科研究的龙头 — 范式，发现了人工智能范式张冠李戴，实施了范式革命，突破了原有人工智能面临的层层障碍，揭示了人工智能的新科学观、新方法论、新模型、新路径、新逻辑、新

数学、新概念、新原理，创建了“机制主义人工智能基础理论”。

具体而言：

● 颠覆了物质学科范式：机械唯物科学观和机械还原方法论

因而

●突破了现行人工智能的全局研究模型：孤立的脑模型

●突破了现行人工智能的研究路径：结构主义、功能主义、行为主义三个分道扬镳的研究路线

●突破了现行人工智能的基础概念：形式数据、形式知识、形式智能

●突破了现行人工智能的逻辑基础：刚性的逻辑

●突破了现行人工智能的数学基础：分立的理论

进而

●确立了信息学科范式：主客互动科学观和信息生态方法论

●构筑了全新的研究模型：主体客体相互作用的演进模型

●开创了全新的研究路径：基于普适性智能生成机制的机制主义路线

●创建了全新的人工智能逻辑理论：泛逻辑理论

●创建了全新的人工智能数学理论：因素空间理论

●重建了人工智能的基本概念：全信息、全知识、全智能

●发现了全新的人工智能基本原理：信息转换与智能创生定律

综之

●首创了《机制主义人工智能基础理论》

简要解释如下。

一切科学研究活动都受到人们的科学观和方法论的引领和支配。科学观阐明了研究对象的本质**“是什么”**，方法论阐明了学科的研究应当**“怎么做”**；科学观和方法论一起，就阐明了学科研究应当遵循的规范方式，简称**“范式”**。因此，**范式是整个学科研究的龙头**，它的正确与否决定了整个学科研究的成败。

显然，作为开放复杂信息系统的人工智能，应当遵循信息学科的范式。但是，我们通过全面调研却惊人地发现：它所遵循的竟然是物质学科的范式，在范式上犯了张冠李戴的大忌！这是现有人工智能一切痼疾顽症的总根源。

比如，在物质学科范式的“分而治之”方法论支配下，人工智能的研究被分解为人工神经网络、专家系统、感知动作系统三个分道扬镳的学派，使人工智能的研究至今仍然处于没有统一理论的初级状态；又如，在物质学科范式的“单纯形式化”方法论支配下，智能的内核（内容要素和价值要素）被完全丢弃，导致现有人工智能系统的结果不可解释，智能水平低下。

为什么人工智能的研究会陷入范式“张冠李戴”的境地？

这是社会法则所决定的。具体来说，由于历史上的科学研究都属于物质学科体系，都遵循物质学科的范式，不存在范式变革的问题。问题发生在 20 世纪中叶以来，信息学科迅猛兴起，形成了信息学科研究的

社会存在。可是信息学科的范式却至今未能形成。这是因为范式属于社会意识范畴，而“**社会意识滞后于社会存在**”是社会发展的根本法则。在这种情形下，信息学科（含人工智能）的研究者们便沿用了当时业已存在的物质学科范式，于是造成了信息学科（含人工智能）范式的张冠李戴，而且一直延续至今，**是为无可避免的结果！**

面对人工智能范式的张冠李戴，唯一有效的解决办法就是实施“正冠”。

我们发现，**中华文明的整体观（科学观）和辨证论（方法论）完全符合信息学科的性质和需要，是引领信息学科和其他复杂科学开拓创新的最佳范式。**据此，我们以中华文明支撑的信息学科范式取代了人工智能的传统学科范式，并在信息学科范式引领下实现了上述各项突破与创新。

成果“机制主义人工智能基础理论”的意义：

●开创了信息学科和复杂学科研究的新范式，具有划时代意义

●从理论上消除了原有人工智能的痼疾顽症，具有重要的理论意义

●智能生成机制是人工智能应用系统的通用孵化平台，具有重大工程意义

●信息转换与智能创生定律与“物质不灭和能量守恒”具有相同科学意义

●见证了中华文明对 21 世纪信息学科和复杂学科开拓创新的引领地位。

654，《机制主义人工智能理论》一书 PPT 简介

《机制主义人工智能理论》
简　介

钟义信
北京邮电大学

zyx@bupt.edu.cn

说　明

由于报告的时间有限，这里给出的是本书的简介。

更详细的内容请参见著作《机制主义人工智能理论》。

目　录

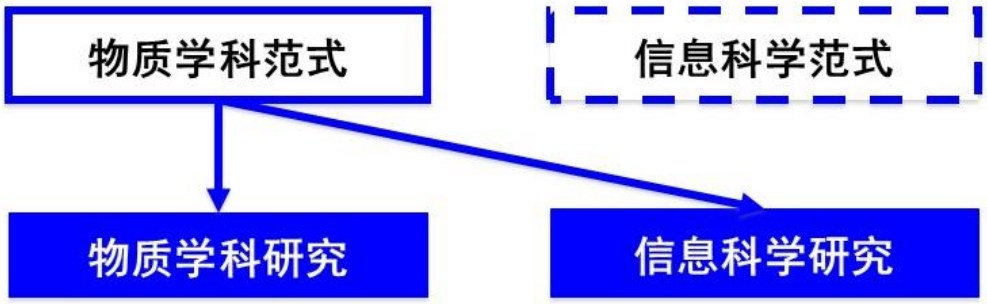
一，观察问题：立足于最高点

—因而发现范式张冠李戴并实施范式革命

发现：范式是学科发展的最高引领者

生长进程	进程名称	进程要素	要素解释
自下而上的摸索阶段	求索（准备）	多方试探	常见：盲人摸象
		总结提炼	任务：寻求范式
自上而下的建构阶段	范式（定义）	科 学 观	宏观回答：是什么
		方 法 论	宏观回答：怎么做
	框架（定位）	全局模型	学科蓝图是什么？
		研究路径	研究路线怎么走？
	规格（定格）	学术结构	交叉结构是什么？
		数理基础	数理方法怎么定？
	理论（定论）	基本概念	基本要素是什么？
		基本原理	内在联系怎么做？

发现：人工智能“范式张冠李戴”是历史的必然



1，物质学科活动及其范式，业已存在数百年（事实）
2，20世纪中叶开始，信息学科研究活动崛起（事实）
3，信息学科范式至今未确立（法则：意识滞后于存在）
4，信息学科借用物质学科范式（法则：思想指导行动）
5，可见，人工智能范式张冠李戴是不可抗拒的规律！

证实：人工智能范式“张冠李戴”

事项	范式要素之一：科学观	范式要素之二：方法论
传统物质科学	物质观 纯客非主，关注结构与功能 确定性演化，可分可合	机械还原方法论 形式化描述，比对式判断 分而治之的全局处置
现行人工智能	准物质观 纯客非主，关注结构与功能 存在不确定性，接受可分性	真还原论 形式化描述，比对式判断 分而治之的全局处置
现代信息科学	信息观 主客互动，关注主客双赢 不确定性演化	信息生态方法论 整体化描述，理解式判断 生态演化的全局处置

人工智能的问题验证了范式张冠李戴

单纯形式化方法论带来的问题（智能被掏空）

- 理解能力低下
- 可解释性很差
- 需要大量样本
- 没有人工情感，没有人工意识

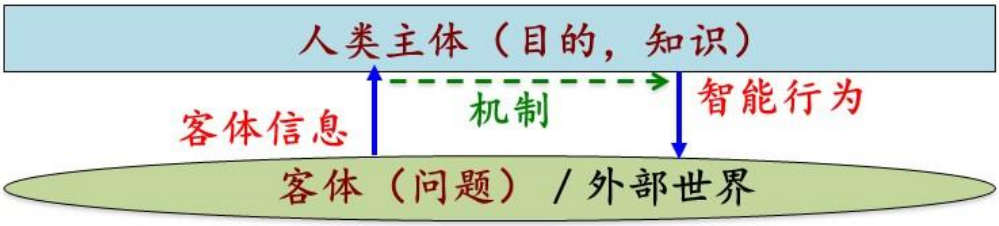
分而治之方法论带来的问题（整体被肢解）

- 学派之间不能实现统一，每个学派内部也难以通用

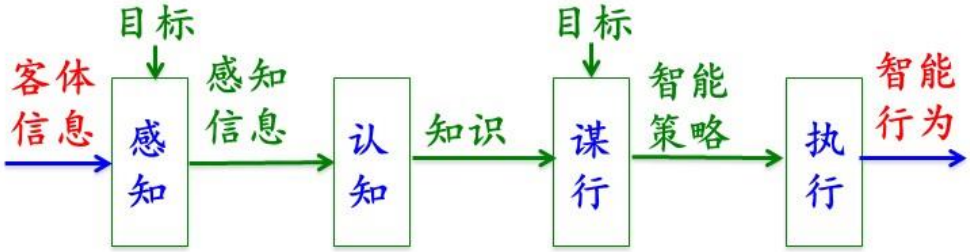
二，求解问题：聚焦于生命线

—因而发现并创建普适性的智能生成机制

智能生成机制是智能系统的生命线



人工智能的模型是主体客体相互作用的信息过程。

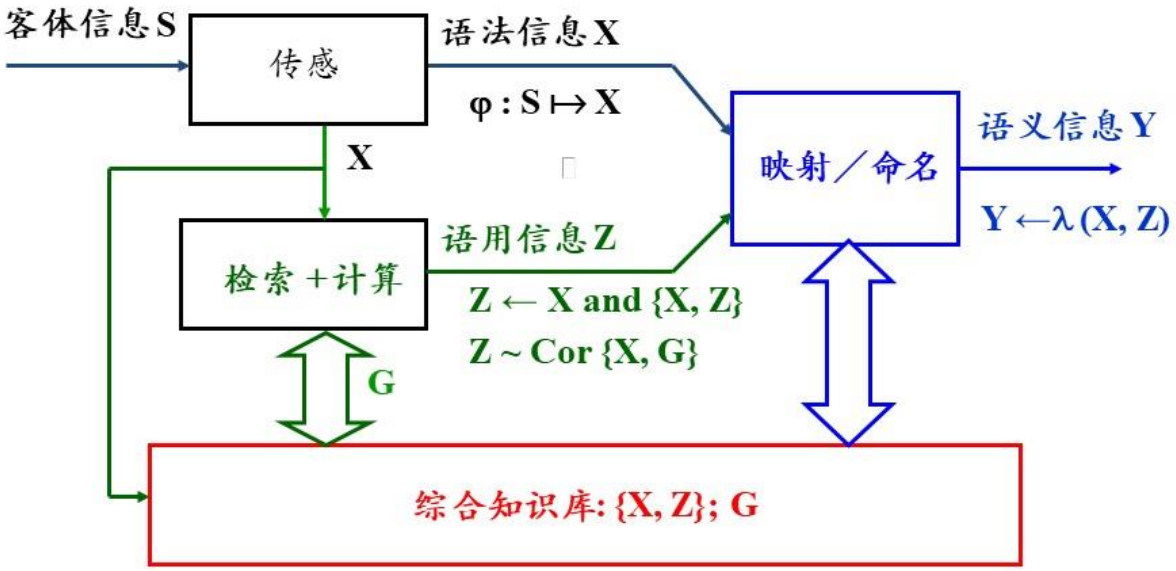


智能生成机制的本质是**信息转换与智能创生原理**。

三，重点突破：着手在最深层

—因而在智能的源头创立“全信息理论”

全信息：智能生成机制的启动者



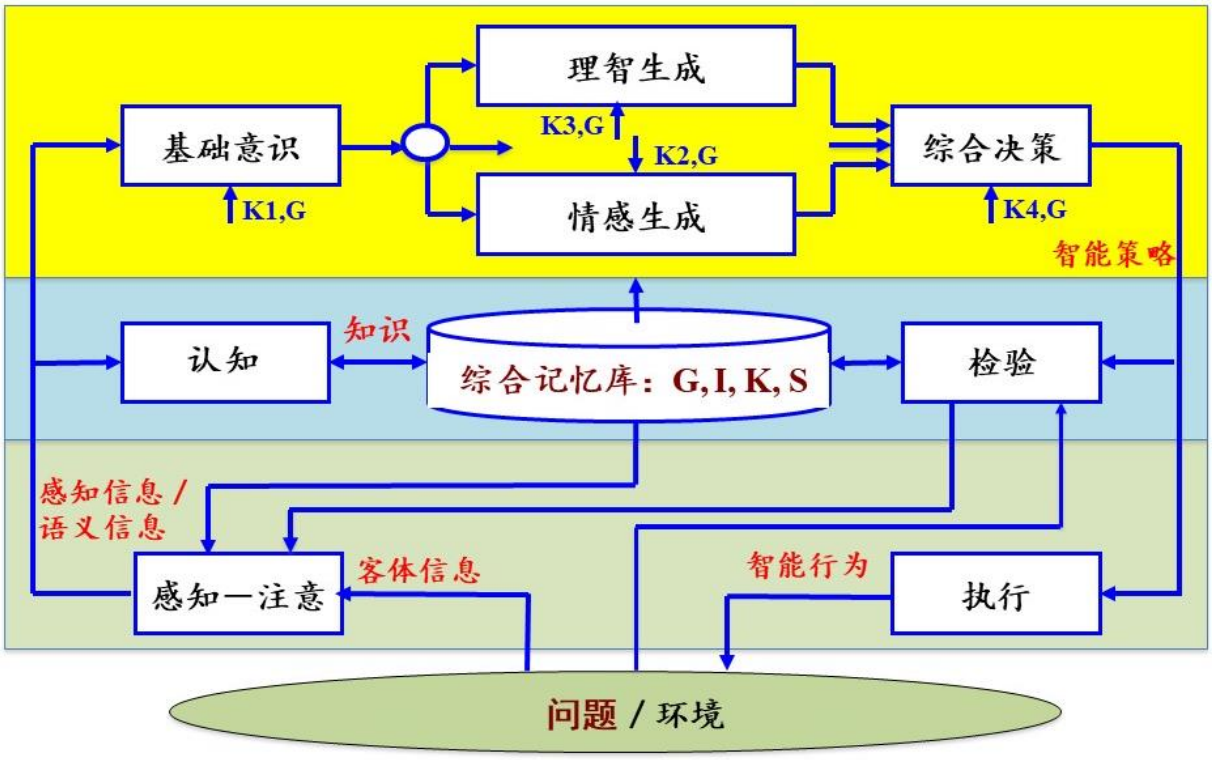
四，系统创新：深耕在无人区

—因而创立“机制主义人工智能理论”

实现了：通用人工智能理论的创生

比较事项	人工智能理论	通用人工智能理论
科 学 观	纯客观，关注结构，确定性	主客互动，关注目的，不确定性
方 法 论	分而治之，单纯形式化	生态演化，形式内容价值一体
全局模型	孤立的脑模型	主客互动的信息过程模型
研究路径	结构、功能、行为	机制（普适性智能生成机制）
学术结构	计算机学科的分支	神经、认知、信息、人文、数理
数理基础	概率论，形式逻辑	泛逻辑，因素空间
基本概念	形式化的数据、知识、智能	全信息、全知识、全智能
基本原理	未总结	信息转换与智能创生定律
基本结果	三个局部理论	通用人工智能理论

基于普适性智能生成机理的通用人工智能基础理论模型



优胜性：全面消除了人工理论的痼疾顽症

颠覆“去主观性”，确立主体的主导性：实现系统的**目的性**

颠覆“分而治之”，创建信息生态方法论，实现**整体性**

突破“脑模型”，创建“主客互动模型”，保障模型**真实性**

突破“三分路径”，揭示“普适性生成机制”，实现**通用性**

突破“纯形式化”，创建了“语义信息论”，奠定**智能基础**

保证充分的“**理解能力**”

保证“**可解释性**”

提供“**小样本学习**”可能性

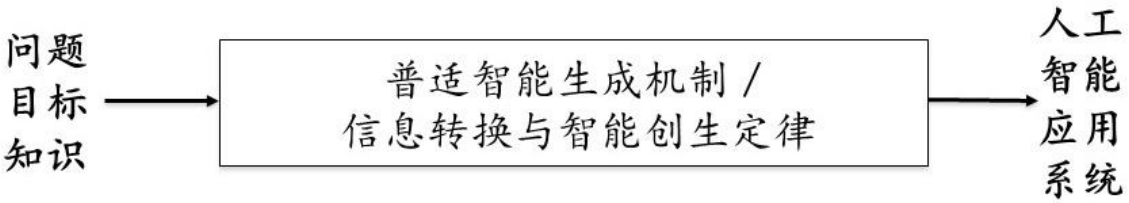
实现“**意识—情感—理智**”和谐生成

五，成果意义：远超人工智能

—确证中华与信息文明在21世纪的优势地位

“信息转换与智能创生定律”的深远意义

（一）工程意义：人工智能孵化平台 — 以不变应万变



（二）学术意义：三大定律

物质领域：质量转换与物质不灭定律

能量领域：能量转换与能量守恒定律

信息领域：信息转换与智能创生定律

范式变革与中华文明在21世纪的优势复归

事项	科学观	方法论
传统物质科学	物质观 纯客非主，关注结构与功能 稳定不变，可分可合	还原论 形式化描述，比对式判断 分而治之的处置
现代信息科学	信息观 主客互动，关注主体目的达成 不确定性贯彻全过程	信息生态方法论 整体化描述，理解式判断 生态演化的处置
中华文明思想	整体观 天人合一，以人为本 易（兵无定势，水无常形）	辩证论 神形兼备，知行合一 道生一，…，三生万物

655，浪潮 AI 服务器已成为全球领先的 AI 算力基础设施平台

据 IDC 报告，2020 上半年全球人工智能（AI）服务器市场规模达 5.9 亿美元，占 AI 基础设施市场的 84.2%以上，成为 AI 基础设施需求的主体。目前浪潮、戴尔、HPE 分列全球 AI 服务器市场份额前三，浪潮占 16.4%，占有率成全球龙头老大。浪潮并连续三年保持中国市场份额 50%以上。

AI 服务器是提供 AI 算力的基础设施的主体，浪潮已成为全球领先的 AI 算力基础设施供应商，布局涵盖训练、推理、边缘等全栈 AI 场景。AI 服务器广泛应用于制造、媒体娱乐、现代农业、智能家居、智慧电力等领域。

656，据保尔森基金会（PaulsonInstitute）麦克罗波洛智库（MacroPolo）公布的“全球 AI 人才追踪”调查，拥有全球人工智能高级研究人员前三的国家，美国占全球的 59%，中国占 10.6%，欧洲占 10.2%。美国人工智能高级研究人员 29%来自中国。

人工智能领域最具代表性的顶级会议 NeurIPS（主要关注神经网络和深度学习方面的理论进展)。前年年底在 NeurIPS2019 大会上，从 1428 篇被接收的论文中抽取 175 篇论文 675 名作者，1/3 研究者来自中国（在中国完成大学本科学习后来在美国公司和大学工作)。中国已经成为全球人工智能研究者最大输出源的国家。2017 年美国国务院把人工智能定义为“引领未来的战略性技术”。争夺人工智能研究人才（特别是高级人才）应是我国制定人才政策的重点。

657，在十四五规划和 2035 远景目标纲要中，列入建设数字中国的重点发展七大产业：云计算、大数据、物联网、VR/AR（虚拟现实/增强现实)、工业互联网、区

块链、人工智能。

658，神经形态（拟态或类脑）芯片算法之优化

北京大学微纳电子研究院黄如院士在第 66 届国际电子器件大会（IEDM）上作神经形态器件报告

北大类脑智能芯片研究中心蔡一茂教授、黄如院士课题组为神经形态计算的器件—阵列—算法协同优化设计提供指导

阻变器件是后摩尔时代构建新型存算一体及类脑芯片、突破冯-诺依曼体系结构瓶颈的关键电子器件技术之一。但阻变器件的非理想效应以及高密度集成带来的热效应会相互耦合，成为阻变器件在存储及神经形态计算应用中的关键挑战。蔡一茂教授、黄如院士课题组，研究阻变器件非理想效应的物理机制，提出了准确描述多种非理想效应的集约模型，建立了能够综合评估器件技术、阵列拓扑及算法设计的跨层次验证平台，掌握了非理想效应和热串扰对存储及神经形态计算应用的影响，为器件-阵列-算法的协同优化设计提供了重要指导。

659，Deepmind 发布人工智能自组装生命框架

SELF-ORGANIZING INTELLIGENT MATTER：A BLUEPRINT FOR AN AI GENERATING ALGORITHM

Karol Gregor，Frederic Besse
DeepMind，UK

摘要：DeepMind 提出新的研究方向，在没有明确智能体概念的情况下，用环境促进智能生物体的出现。

近日,DeepMind 的研究者提出了一种人工生命框架,旨在促进智能生物体的出现。该框架中没有明确的智能体概念，而是由原子元素构成的环境。这些元素包含神

经操作，通过信息交换和环境中包含的类物理规则进行交互。研究者讨论了进化过程如何导致由许多此类原子元素构成的不同生物体的出现，这些原子元素可以在环境中共存和繁荣。此外，研究者还探讨了这如何构成通用 AI 生成算法的基础，并提供了这种系统的简化版实现，讨论了需要做哪些改进才能进一步扩大规模。

全文：DeepMind 提出新系统

现实世界是由相互作用并组成更大实体的基本粒子构成的。DeepMind 研究提出的环境（AI 生成算法）是由元素构成的，但尺度较大。每个元素包含一个神经操作，比如矩阵乘法、外积，或者是包含这些算子的序列。这些元素通过某种形式的基本规则——一种物理类型，以及神经状态的直接通信进行彼此交互。

该系统有多种实现。这篇论文提供了网格世界（grid-world）实现，其中的基本元素位于网格上，通过传播信号或注意力机制进行通信，并与实现能量和类化学交换的底层物理进行通信。另一个例子是在三维空间中形成刚性零件的元素，这些零件可以通过连接点（joint）进行连接，连接点包含神经操作，通过与附近连接的零件交换信号来进行交互，并在连接点上设置扭矩。系统中可能存在多种类型的元素，并非所有元素内部都需要有神经网络。

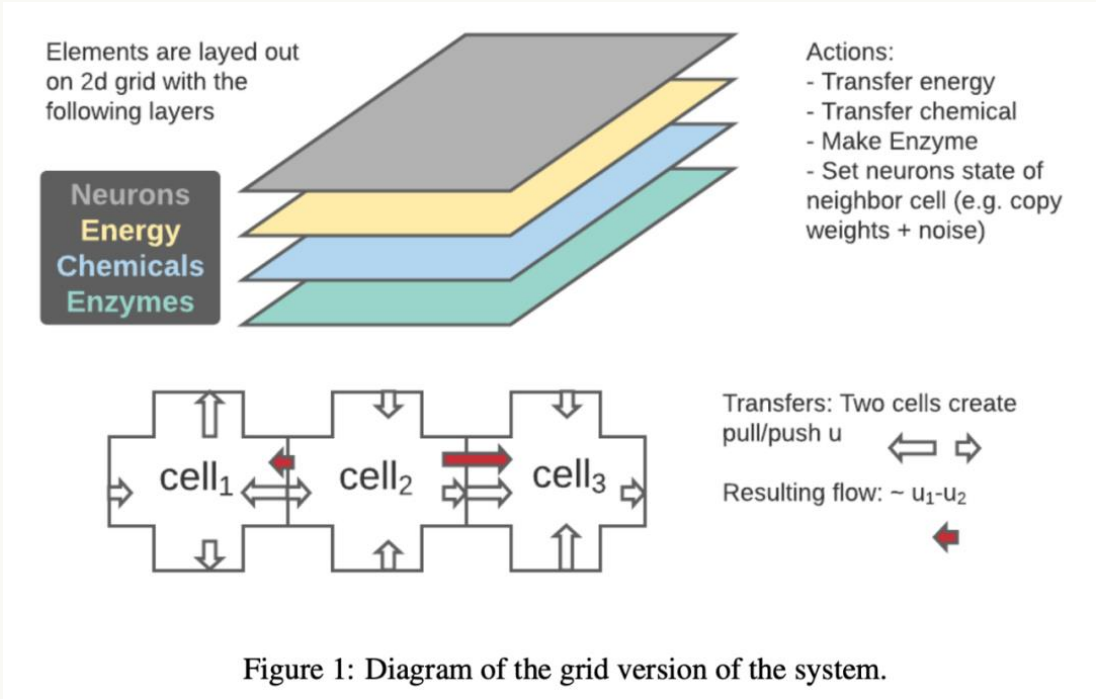


Figure 1: Diagram of the grid version of the system.

研究者在论文中提供了一种网格实现，突出显示了许多重要属性，并探讨了要让该系统变得强大需要进行哪些改进。

但是，该系统的潜力是无限的，它支持如下功能：

由多个元素组成的较大单元可以通过物理连接（如机器人）来形成，也可以简单地作为一组决定进行通信并形成整体的单元。这些单元的潜在大小没有限制。它们可以通过多种方式传播——通过接管环境中的其他元素来生长（殖民地），也可以通过组装新的副本进行复制，将适当的收集元素移动位置（例如机器人通过组装碎片来复制自己）或自我组装，或者它们可以生成完全不同的单元，这些单元可以实现专用的功能（一种有用的机器），或者比其前代产品更好的单元。而后者可能需要智能（intelligence）。

智能的能力

为什么说该研究提出的计算系统具有表示通用智能的能力，研究者提供了两个论点：

首先，机器学习中已有的任何神经算法，或者未来可能创建的算法，都可以写作一串操作序列，例如加法、矩阵乘法、外积和非线性运算，并在张量状态下进行操作，例如由神经网络的前向、后向和优化器操作产生的序列。AutoML-Zero 意识到了这一点，它直接搜索此类算子的序列以及与其所运行状态的连通性，并且能够学习基本的神经算法。由于这些算子是环境的基本构建元素，且能与任意连通性进行通信，因此所有的神经算法都可以在该系统中实现。

智能体假设

在该系统中，没有智能体和环境之分，只有环境。元素本身可能形成也可能不形成进化单元，进化单元的繁殖会显示出遗传性但遗传的区域并不确切。在前一种情况下，它们可以自主移动，收集能量并进行复制，形成更大的聚集体或复制生物体，因为这样做具有优势。而该研究则是针对后者，它需要最小数量的更简单协作单元进行自我传播。

SIM 的网格版本和通用属性

该论文还介绍了自组织智能物质（self-organizing intelligent matter，SIM）的实例，讨论了其各个方面，并提供了该研究认为它能够构成 AI 生成算法的更多原因。

如上文所述，这里没有内置的智能体概念，实际上只有一个环境。通常情况下，在两个不同平台上实现该系统是很不自然的：一个用于物理部分，例如物理模拟器；一个用于神经部分，如 TensorFlow、PyTorch 或 Jax 等神经网络框架。该研究建议在单个平台上制作这样的系统。要产生智能行为，需要高效地运行神经网络，因此该系统需要在后一种平台上实现。出于灵活性的考虑，该研究选择了 Jax。

Jax 在张量上运行，该研究用张量来存储元素。这些元素需要交互，并具有形成任意大小灵活聚合体的能力。

实验

该研究运行了上述系统，在一系列运行之间观察到了令人兴奋的多样性，如图 2 所示。

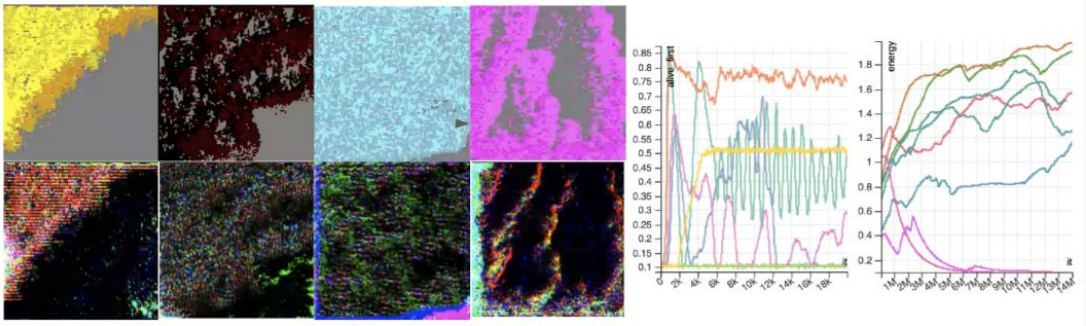


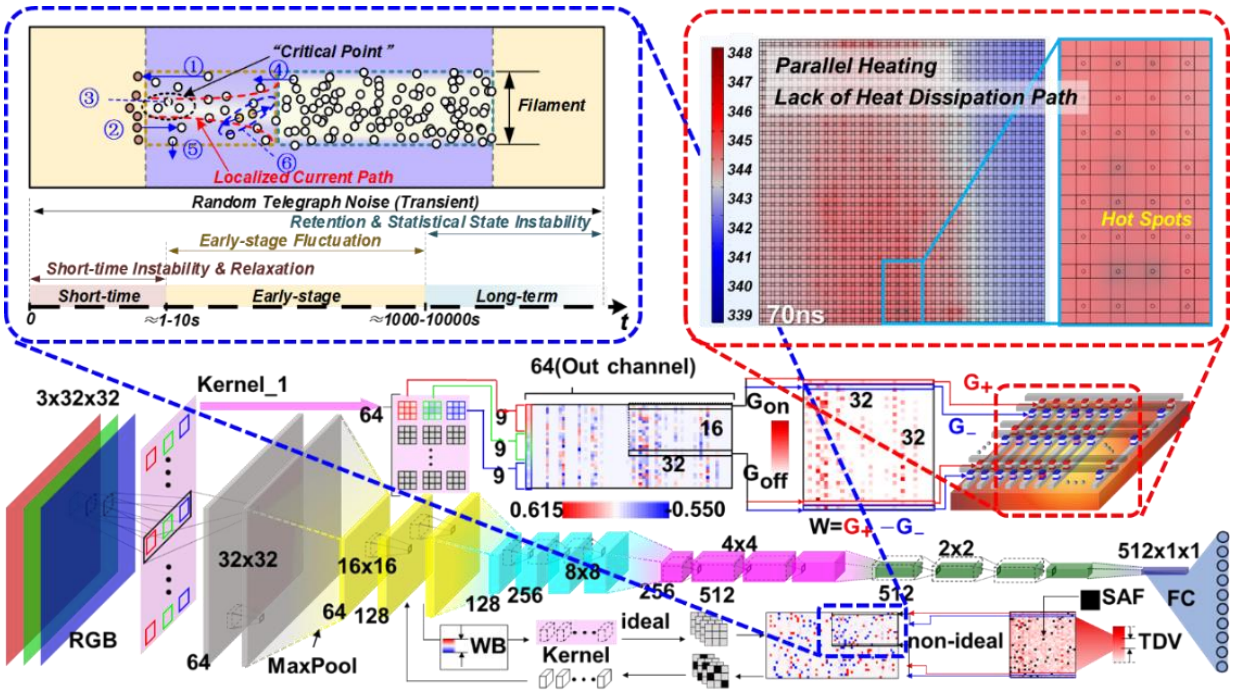
图 2：运行结果。上面一行中，研究者用不同颜色代表 3 种不同的随机权重。

如图 2 上面一行中我们可以看到，在多个区域中，两种元素都能够稳定共存，即相同空间区域中存在不同颜色的点。并且这能持续很长时间，说明它们发现了一种共存的方式。

660，阻变器件——存算一体的类脑芯片

北京大学黄如院士研发团队

北大微纳电子研究院黄如院士在第 66 届国际电子器件大会（IEDM）上（2020.12.12-18）上发表他们研发的神经形态器件（类脑智能芯片）的论文。阻变器件是后摩尔时代构建新型存算一体及类脑芯片、突破冯-诺依学体系结构瓶颈的关键电子器件技术之一。但阻变器件的非理想效应以及高密度集成带来的热效应会相互耦合，成为阻变器件在存储及神经形态计算应用 1 中的关键挑战。蔡一茂教授、黄如院士团队系统研究了阻变器件非理想效应的物理机制，提出了准确描述多种非理想效应的集约模型，建立了能够综合评估器件技术、阵列拓扑及算法设计的跨层次验证平台，掌握了非现想效应和热串扰对存储及神经形态计算应用的影响，为器件-阵列-算法的协同优化设计提供了重要指导。



图，阻变存储技术在存储和神经形态计算应用中的器件-阵列-算法协同优化

661，关于隐马尔可夫模型混合的可解释性

Towards interpretability of Mixtures of Hidden Markov Models

Negar Safinianaini,Henrik Boström

瑞典斯德哥尔摩 KTH 皇家技术学院　2021.3.23

隐性马尔可夫模型（MHMM）的混合通常用于顺序数据的聚类。与任何聚类方法一样,MHMM 的一个重要方面是它们的可解释性,从而可以从数据中获得新颖的见解。但是如果没有适当的衡量可解释性的方法，那么对新颖贡献的评估就很困难，并且实际上不可能设计出直接优化此特性的技术。在这项工作中，提出了一种用于 MHMM 的可解释性的信息理论测度（熵），并在此基础上找出了一种改进模型可解释性的新方法，即一种熵正则化期望最大化（EM）算法。新方法旨在减少 MHMM 中的马尔可夫链（涉及状态转移矩阵）的熵，即在聚类期间为常见状态转移分配更高的权重。有人认为，这种熵的减少通常会导致可解释性的提高，因为可以更容易地识别出群集中最有影响力和最重要的状态转换。一项实证研究表明，可以通

过熵来改进 MHMM 的可解释性，而不必牺牲（但要提高）聚类性能和计算成本（分别通过 v 度量和 EM 迭代次数来度量）。

662，带有自适应脉冲递归神经网络的准确高效时域分类

Accurate and efficient time-domain classification with adaptive spiking recurrent neural networks

Bojian Yin, Federico Corradi, Sander M. Bohte
荷兰阿姆斯特丹大学

（荷兰阿姆斯特丹大学在 arxiv 平台上发表一篇将脉冲神经网络应用到时域相关问题的文章）

受更详细的生物神经元建模启发，脉冲神经网络（SNN）已被研究为生物学上更合理，潜在能力更强大的神经计算模型；但是，与传统的人工神经网络（ANN）相比，此类网络的性能仍然不足。在这里，我们在时域中具有挑战性的基准问题（语音和手势识别）领域，提出了 SNN 的最新技术—将最新的替代梯度与可调谐和自适应脉冲神经元的递归网络结合。这一技术超出了标准经典递归神经网络（RNN）的性能，并且接近最佳现代 ANN 的性能。由于这些 SNN 表现出稀疏的脉冲，因此我们证明，与具有可比性能的 RNN 相比，它们在理论上的计算效率高出 1-3 个数量级。总之，这可以将 SNN 定位为 AI 硬件实现的有吸引力的解决方案。

663，英国急诊医学临床医生采用基于机器学习（ML）的临床决策支持系统（CDSS）

今天由于医疗费用飞涨、患者对临床服务需求增长、收集到的患者数据激增和越来越复杂的患者诊治的选择，医疗资源分配受到高度重视。

在英国，医务人员经常求助于临床决策支持系统（CDSS）的工具来辅助决等过程。

在今天大数据时代，除了收集越来越多的患者信息外，CDSS 数量不断增加。人们对基于 ML（人工智能子集）支持的 CDSS 在医疗决策中的应用尤感兴趣！

ML 运用算法和统计模型，在没有明确人类指令的情况下执行任务。ML 算法通过模式识别和从自已过去的数据中推断来训练 ML 工具，从而对未来的新数据做出最佳的预测或前瞻性建议。

664，基于新的多任务基准 RAINBOW 的通用常识推理模型 UNICORN

Nicholas Lourie, Ronan Le Bras, Chandra Bhagavatula, Yejin Choi
美国艾伦人工智能研究所、华盛顿计算机科学与工程学院

https://arxiv.org/abs/2103.13009

长期以来，常识 AI 一直被视为几乎不可能实现的目标。现在随着新的基准和模型的涌入，对此研究的兴趣急剧增加。

本文提出两种评估常识模型的新方法，强调它们在新任务上的通用性，并在最近引入的各种基准上建立基础。本文提出一个新任务基准 RAINBOW，以促进对常识模型的研究，这些常识模型可以很好地概括多个任务和数据集。其次本文提出了一种新颖的评估方法，即成本当量曲线，它对源数据集，预训练的语言模型和转移学习方法的选择如何影响性能和数据效率给出了新的见解。

作者进行了广泛实验(超过 200 个),涵盖 4800 个模型,并报告了许多有价值的、有时令人惊讶的发现。例如，如果遵循特定的方法，迁移几乎总是会带来更好或同等的性能，基于 QA 的常识数据集之间的迁移很好；而常识知识图谱则不然，而且可能与直觉相反，较大的模型从迁移中受益更多。

最后作者引入了一种新的通用常识推理模型 UNICORN，该模型在 8 种流行的常识基准上得到了最优的性能,分别是 aNLI(87.3%),CosmosQA(91.8%),HellaSWAG

（93.9%），PIQA（90.1%），SocialIQa（83.2%），WinoGrande（86.6%），CycIC（94.0%）和 CommonsenseQA（79.3%）。

665，商汤科技研发的人脸识别在全球领先

德国《商报》在 2021.4.3.一篇“中国对全球经济攻势”的文章中谈到商汤科技（Sensetlime）的人脸识别一体机，0.3 秒极速验证准确率 99.99%。2 月下旬，谷歌前 CEO 埃里克-施密特受拜顿总统委托，在评估中美人工智能发展水平时也谈到：“在人脸识别上，中国超过了美国，在全球是顶尖的”。

666，关于图神经网络的可解释性

Hao Yuan, Haiyang Yu, Shurui Gui, and Shuiwang Ji

华盛顿州立大学，2021.3.25

深度学习方法在许多人工智能任务中的表现越来越突出。深度模型的一个主要局限性是它们不易解释。这一限制可以通过研究解释技术来规避，从而产生了可解释性领域。近年来，图像和文本深层模型的可解释性取得了重大进展。在图形数据领域，图神经网络及其可解释性得到了迅速发展。本文提出对当前 GNN 解释方法的分类观点，阐明了现有方法的共性和差异，为进一步的方法发展奠定了基础。为了方便评估，本文生成了一组用于 GNN 解释性的基准图数据集。本文还总结了当前用于评估 GNN 解释性的数据集和度量。总之，本文为 GNN 解释性提供了一个统一的方法论处理和一个标准化的评估试验台。

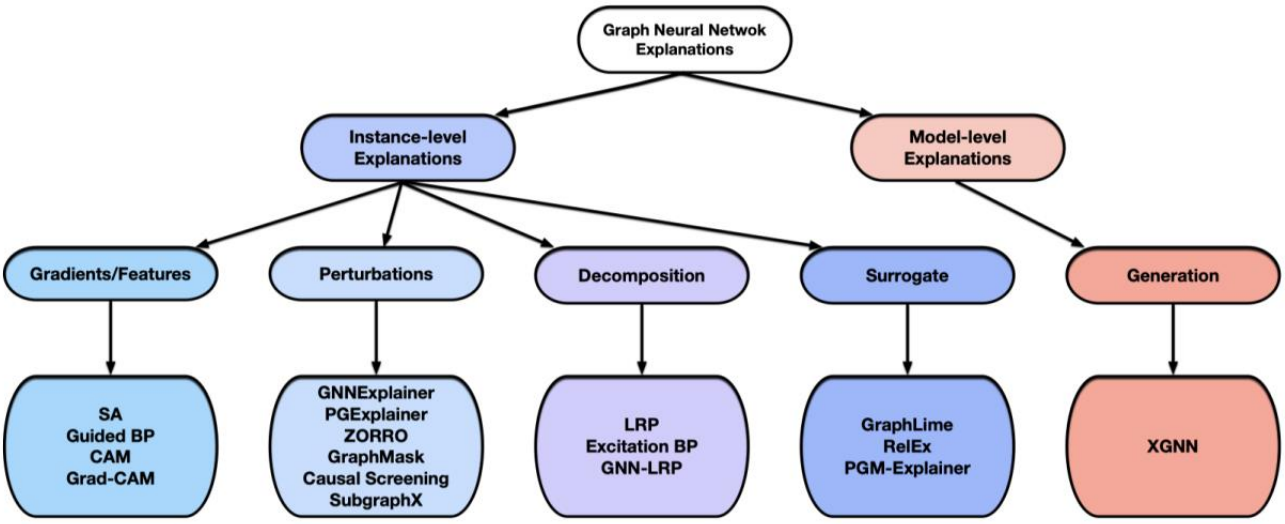
对图神经网络可解释方法的分类如下附图所示，分为两大类：基于实例的方法和基于模型的方法。

基于实例的方法为每个输入图提供依赖于输入的解释。给定一个输入图，这些方法通过识别用于预测的重要输入特征来解释深层模型。基于实例的方法又可分为四个不同的分支：基于梯度/特征的方法，基于扰动的方法，基于分解的方法和基于替代的方法。具体而言，基于梯度/特征的方法使用梯度或特征值来表示不同输入特性的重要性。基于扰动的方法可以观察预测结果在不同输入扰动下的变化，以研究输入重要性得分。基于分解的方法首先将预测分数（如预测概率）分解到最后一个隐藏层的神经元，然后将这些分数逐层反向传播到输入空间，并将分解后的分数作为重要度分数。

基于模型的方法在解释图神经网络时，与输入无关的解释是高层次的，可以解释一般行为。唯一现有的基于模型的方法是基于图生成的 XGNN 方法。它生成图模式来最大化某个类的预测概率，并使用这些图模式来解释这个类。

总的来说，这两类方法从不同角度解释了图模型。基于实例的方法提供了对特定实例的解释，而基于模型的方法提供了对图模型如何工作的一般理解。对于实例的方法，为了验证和信任图模型，需要人对不同输入实例的解释进行验证。例如由于专家需要探索不同输入图的解释，因此需要更多的人为监督。对于模型级方法，由于解释是高层次的，因此较少涉及人为监督。此外，实例级方法的解释是基于实际输入实例的，因此它们很容易理解。然而，对于模型级方法的解释可能是人类无法理解的，因为获得的图模式甚至可能不存在于现实世界中。总的来说，这两种方法可以结合在一起，以便更好地理解图模型。

附图：



667，论系统软件在神经拟态计算能量管理中的作用

Twisha Titirsha, Shihao Song, Adarsha Balaji, Anup Das

德雷克塞尔大学，2021.3.22

最近，诸如 DYNAP 和 Loihi 的神经拟态计算系统被引入计算机界，以提高机器学习程序的性能和能效，尤其是那些使用 Spiking Neural Network（SNN）实现的机器学习程序。用于神经拟态系统的系统软件的作用是将大型机器学习模型（例如具有许多神经元和突触）聚类并将这些聚类映射到硬件的计算资源。本文考虑了神经元和突触所消耗的功率，以及在互连上传递尖峰时所消耗的能量，从而制定了神经拟态硬件的能量消耗。

基于这样的表述，本文首先评估系统软件在管理神经拟态系统能量消耗中的作用。采用一种简单的基于启发式的映射方法，将神经元和突触放置到计算资源上以减少能耗。本文还用 10 个机器学习应用程序评估了本文提出的方法，并证明了所提出的映射方法可显著减少神经拟态计算系统的能耗。

668，使用脉冲间隔对脉冲神经网络进行视觉解释

Youngeun Kim, Priyadarshini Panda

耶鲁大学，2021.3.26

脉冲神经网络（SNN）可以与异步二进制时间事件进行计算并进行通信，这可以通过使用神经形态硬件来大大节省能源，最近 SNN 相关算法工作已显示出在各种分类任务上的良好性能。然而，目前缺乏研究用于分析和解释这种深度 SNN 的内部脉冲行为的可视化工具。在本文中，我们提出了一种针对 SNN 的生物可视化的新概念，称为峰值激活图（SAM）。拟议中的 SAM 通过消除计算梯度以获得视觉解释的需要，规避了脉冲神经元的不可微特征，相反，SAM 通过在不同时间步长上正向传播输入脉冲来计算时间可视化图。SAM 通过突出显示具有短峰间间隔话动的神经元来产生与输入数据的每个时间步相对应的注意力图。有趣的是，在没有反向传播过程和类标签的情况下，SAM 会突出显示图像的区分区域，同时捕获细粒度的细节。借助 SAM，我们首次根据优化类型，泄漏行为以及何时遇到对抗性示例，对内部脉冲在各种 SNN 训练配置中的工作方式进行了全面分析。

669，用多尺度稀疏卷积网络图表对帕金森病进行步态评估

上海交大生物医学工程学院钱晓华课题组、上海交大医学院瑞金医院孙伯民团队、张陈诚医生

《IEEE Transactions on Multimedia》（TMM, JCR Multimedia No.1, IF:6.05）发表

帕金森病是一种渐进式的神经退行性疾病，运动障碍是 PD 最典型的临床特征和最易识别的症状之一。目前帕金森病运动功能评估的主要依据是使用临床评分量表（即 MDS-UPDRS）。临床实践中对该量表的评定方式存在两个弊端：①这种评估方式费时费力，并且临床评估医生经验水平的差异常常导致评估结果存在较大的

主观性；②不利于实现帕金森病患者移动化的家庭管理和及时的病情反馈。如何自动而客观地进行帕金森病患者的运动功能评估已成为当前重要的研究热点。

步态运动障碍是帕金森病中常见的运动障碍之一，主要通过对患者的步幅大小、步速、足部离地高度、走路时足跟着地的情况、转身和摆臂等临床分析来评估。步态评估与帕金森病的严重程度密切有关，也是 MDS-UPDRS 中运动功能检查的重要组成部分。因此，步态运动障碍的自动量化评估对实现帕金森病患者的自动运动功能评估至关重要。

作者开发了一种多尺度稀疏化时空卷积网络模型来实现帕金森病患者步态运动障碍的自动定量评估。具体来说：

①提出了双流时空图卷积网络，对从视频中提取的人体关节序列和对应的骨架序列进行时空建模；

②建立了深度监督下的多尺度时空注意力感知机制，以在不同尺度之间的强相关性下捕获多尺度细粒度时空特征；

③提出了一种模型驱动的稀疏化策略，实现判别性特征的选择。

该模型在上海交大医学院附属瑞金医院功能神经外科提供的临床视频数据集上进行了全面的评估，该数据集包含 2017～2019 年 142 位患者的 441 个步态视频。这也是目前帕金森病步态自动评估研究中最大的数据集，最终该模型实现了 65.66%的准确率和 98.90%的可接收准确率，优于其他已有的帕金森病步态自动评估方法。该论文提出的非接触式方法为帕金森病患者步态视频的自动定量评估提供了一种新的潜在工具。

670，得益于人工智能提高核聚变反应堆计算速度

利用机器学习的方法，可取得核聚变反应堆堆芯等离子体湍流传输的数据，进行建模(数值模型／Nu medical models),为加快反应堆设计和有效运行创造条件，以经济可行的方式实现核聚变反应堆净功率的提升。

核聚变反应堆技术可以安全和可持续的方式满足世界未来的电力需求。

Aaron Ho，2021.3.17

埃因霍温理工大学应用物理系核聚变科学与技术组

671，探讨如何利用脑电技术提供人工耳蜗植入儿童听觉康复评估方法

设计适合低龄儿童人工耳蜗的实验范式，为脑机接口在儿童听觉康复方面的应用奠定基础，有助于为人工耳蜗调试和听觉言语康复训练提供更准确的参考依据。

本论文由国际著名听力杂志《Hearing Research》2020.12.27 发表

天津大学神经工程倪广健、刘琪研发团队及明东教授(通讯),与国家儿童医学中心、首都医科大学附属儿童医院刘海江教授合作

672，人脸伪造检测技术

（基于单中心损失监督的频率感知鉴别特征学习框架）

中科大、快手 2021.3.16

随着基于自编码器和生成对抗网络的图像生成技术的快速发展，以 Deep Fake 为代表的人脸伪造技术在娱乐大众的同时，也带来巨大的安全隐患。与之对应的，人脸伪造检测也逐渐成为计算机视觉领域研究的热点。来自中科大、快手的研究者针对人脸伪造，提出了基于单中心损失监督的频率感知鉴别特征学习框架，将度量学习和自适应频率特征学习应用人脸伪造检测，实现 SOTA 性能。

673，微米机器人首次实现小鼠脑瘤的主动靶向治疗

Hongyue Zhang、Zesheng Li、Changyong Gao、Xinjian Fan、Yuxin Pang、Tianlong Li

哈尔滨工业大学，2021.3.24

游动微型机器人(例如细菌或精子驱动的微型机器人)具有自我推进和导航能力，由于其在身体内难以触及的区域运动，用于非侵入性药物递送和治疗，已成为一个令人兴奋的研究领域。然而，目前基于细胞的微型机器人在进入人体后容易受到免疫攻击。日前，哈工大微纳米技术研究中心首次实现游动微米机器人对脑胶质瘤的主动靶向治疗，该技术可以实现在身体内主动地将药物运送到恶性胶质瘤，可抑制肿瘤细胞的增殖。对于肿瘤和癌症而言，靶向治疗是最有效的方法。靶向治疗对于脑瘤而言更具挑战性，这是因为脑部手术非常危险。微型机器人或许提供了一种更安全的治疗手段。

674，解释人工智能所做决策的工作手册

David Leslie,Morgan Briggs，2021.3.20

英国阿兰图灵研究所

https：https://arxiv.org/pdf/2104.03906.pdf

本手册由信息专员办公室和阿兰图灵研究所共同编制，概述了如何应用人工智能可解释性原则和实践。本手册介绍了解释人工智能决策的基础知识，提供了人工智能可解释性的四个原则、人工智能解释的类型，以及人工智能／多语言系统的解释性设计、开发和使用所涉及的任务。

人工智能可解释性的四个原则：

①透明。透明原则是 GDPR（合法性、公平性和透明度）中原则（a）的透明方向

的延伸。在数据保护方面，透明度意味着对你是谁以及如何和为什么使用个人数据保持开发和诚实的态度。人工智能辅助决策的透明性建立在这些需求之上。它让你使用人工智能做决策的过程变得透明，并以一种有意义的方式向个人解释你所做的决定。

②负责。该原则源于 GDPR 中的责任制。在数据保护术语中，责任意味着承担遵守数据保护原则的责任，并能够证明遵守这些原则。负责解释人工智能辅助的决策将这些要求集中在设计和部署人工智能模型时执行的过程和操作上。

③考虑上下文。解释人工智能辅助决策不可以一刀切，需要关注几个不同但互相关联的元素，这些元素可以对解释人工智能辅助决策和管理整个过程产生影响。从概念到部署，以及向决策接受者介绍解释的各个阶段都需要考虑该原则。

④反思影响。在做出决策和执行任务之前需要有要负责的人去思考和推理，人工智能系统越来越多地充当人类决策的受托人。然而我们不能让这些系统直接对其结果和行为负责。在整个开发和实施阶段，你应该重新审视并反思 AI 项目初始阶段确定的影响。如果发现了任何新的影响，你至少应该记录这些影响，并思考如何减轻这些影响。这将帮助你向决策接受者解释你已确定的影响以及你如何尽可能减少任何潜在的有害影响。

解释人工智能所做决定的方式很多，现列出 6 种解释类型和说明，并对每种解释类型进行详细描述。

6 种解释类型和说明：

①基本原理解释

人工智能系统做决策的原因

②责任说明

谁参与了人工智能系统的开发、管理和实施，以及谁对决策进行人工审查

③数据说明

在特定决策中使用了哪些数据以及如何使用这些数据

④公平性解释

在人工智能系统的没计和实施过程中，为了确保其支持的决策是无偏见和公平的，所采取的步骤

⑤安全和性能说明

在人工智能系统的设计和实施过程中，为了最大限度地提高决策和行为的准确性、可靠性、安全性和稳健性，所采取的步骤

⑥影响解释

在人工智能系统的设计和实施过程中，为了考虑和监控人工智能系统的使用及其决策对个人和社会的影响，所采取的步骤

675，图像分类任务中卷积神经网络解释的白盒方法

Meghna P Ayyar,Jenny Benois-Pineau,Akka Zemmari

法国 Bordeaux 大学，LaBRI Crsdela 实验室，2021.4.6.

https://arxiv.org/pdf/2104.02548.pdf

近年来，深度学习已成为解决来自多个领域的应用程序的普遍方法。卷积神经网络（CNN）特别展示了用于图像分类任务的最新技术性能，但是这些网络做出的决定并不透明，不能由人直接解释，已经提出了几种方法来解释以理解网络做出的预测背后的原因。在本文中，提出了一种基于这些方法的假设和实现对这些方法进行分组的拓扑。作者主要关注白盒方法，这些方法利用网络内部体系结构的信息来解释其决策。给定图像分类和受过训练的 CNN 的任务，这项工作旨在提供一

套全面而详细的方法概述，该方法可用于为特定图像创建解释图，这些解释图为图像的每个像素分配重要性得分基于其对网络决策的贡献。作者还建议根控其实现方式对白盒方法进行进一步分类，以实现更好的比较并帮助研究人员找到最适合不同情况的方法。

676，利用神经拟态网络预测癫痫的全新算法

Fengshi Tian, Jie Yang, Shiqi Zhao, Mohamad Sawan

CenBRAIN 实验室

现有的较为有效的癫痫发作预测方法多为卷积神经网络（CNN）算法，他们都具有特异性和敏感性。但是 CNN 在计算上很昂贵并且耗电。这些不便之处使基于 CNN 的方法难以在可穿戴设备上实现。本文受高能效尖峰神经网络（SNN）的启发，在这项工作中提出了一种全新的用于预测癫痫发作的神经拟态算法。这种方法使用设计的高斯隨机离散编码器从 EEG 样本生成尖峰序列，并在结合了 CNN 和 SNN 优势的尖峰卷积神经网络（Spiking-CNN）中进行预测。实验结果表明，与 CNN 比，灵敏度、特异性和 A∪C 分别可得持 95.1%、99.2%和 0.912，计算复杂度降低了 98.58%，表明所提出的 Spiking-CNN 具有硬件友好性和高精度。

677，机器学习技术帮助瘫痪患者进入大脑自然学习过程，从而解除采用脑机接口患者每天都要对该系统重新设置和校正

加州大学旧金山分校的研究人员表明，机器学习技术帮助瘫痪的个体通过他们的大脑活动来学习控制计算机光标，而无需每天进行大量的再训练。

近年来，BCI 领域取得很大进步，但是由于必须每天对现有系统进行重置和重新校正，因此他们无法利用大脑的自然学习过程。加州大学旧金山分校资深研究学

者、神经病学副教授 Karunesh Ganguly 博士说“这就像要有人从头开始学习一遍又一遍地骑自行车”。大脑复杂的长期学习模式可以使人平稳地工作，这是瘫痪者从未有过的表现。

脑机接口领域近年来取得了进步，但由于现有的系统每天都要重新设置和校准，它们还不能进入大脑的自然学习过程。让人工学习系统适应大脑复杂的长期学习模式，这是以前从未在瘫痪患者身上展示过的。

678，国内研制 RISC-Ⅴ DSP 芯片即将量产突破美国芯片封锁

2021.4.12 讯，中科昊芯公司研制的 RISC-Ⅴ DSP 芯片即将量产，突破美国芯片封锁。

RISC-Ⅴ是与 X86、ARM 并列为三大指令集架构，RISC-Ⅴ是开源的，DSP 芯片可对数据进行实时处理，为华为的鸿蒙系统推广提供平台。

679，多跳推理真的可以解释吗？走向基准推理的可解释性

Xin Lv1,2 , Yixin Cao3 , Lei Hou1,2 , Juanzi Li1,2 Zhiyuan Liu1,2 , Yichi Zhang4 , Zelin Dai4

清华大学，南洋理工大学，阿里巴巴

近年来，多跳推理被广泛研究以获得更多可解释的链接预测。然而，我们在实验室中发现，这些模型给出的许多路径实际上是不合理的，而对它们的可解释性评估却做得很少。本文提出了一个统一的框架来定量评估多跳推理模型的可解释性，并设计了一个近似策略来使用规则的可解释性得分来计算它们。此外，我们手动注释了所有可能的规则，并建立一个基准来检测多跳推理的可解释性（BIMR）。在实验中，我们在基准上运行了 9 个基准。实验结果表明，当前多跳推理模型的可解释性不太令人满意，并且仍然远低于我们的基准所给出的上限。此外，基于规

则的模型在性能和可解释性方面优于多跳推理模型，这为未来的研究指明了方向，即我们应该研究如何更好地将规则信号纳入多跳推理模型。

680，随机森林的结论性局部解释规则

Ioannis Mollas,Nick Bassiliades,Grigorios Tsoumakas

Aristotle 大学，2021.4.15

在涉及歧视、性别不平等、经济损失，甚至可能造成人员伤亡的严峻形势下，机器学习模型必须能够为决策者的决策提供清晰的解释。否则，他们晦涩的决策过程干扰人们的生活，从而导致社会道德问题。隨机森林算法一直在努力发展，因为其自解释能力是显而易见的要求。本文介绍 Lion Forests，它是一种特定于森林的隨机解释技术，提供规则作为解释。它适用于从二元分类任务到多元分类和回归任务，并且具有稳定的理论背景。它还进行了实验，包括敏感性分析和最新技术的比较。最后，本文还重点介绍 Lion Forests 的一个独特属性，即结论性，它提供解释有效性，并将其与以前的技术区分开。

681，基于异构网络和混合编码的快速智能神经拟态传感器

Angel Yanguas-Gil

阿贡国家实验室 2021.4.9

本文以昆虫的大脑为模型，了解如何利用异构结构（结合不同类型的神经元和编码）来创建集成输入处理、评估和响应的系统。本文还展示了如何利用时间和速率编码的组合构建高速传感器，该传感器能够在几个周期内基于输入生成假设，然后将该假设作用辅助输入以进行更详细的分析。

https://arxiv.org/pdf/2104.04121.pdf

682，QA 一 GNN：基于语言模型和知识图谱的问答推理

Michihiro Yasunaga,Hongyu Ren,Antoine Bosselut,Percy Liang Jure Leskovec

斯坦福大学（Stanford）

目前关于使用来自事先训练的语言模型（LM）和知识图谱（KG）的知识来回答问题的难题还存在两个挑战：给定 QA 背景（问题和答案的选择），方法需要①从大型 KG 中识别相关知识，②在质量检查环境和 KG 中执行联合推理。在这里，作者们提出了一种新的模型 QA-GNN，它通过两项关键创新解决了上述挑战：①相关性评分，其中作者们使用 LM 来估计 KG 节点相对于给定 QA 上下文的重要性，以及②联合推理，作者们将 QA 上下文和 KG 连接起来以形成一个联合图，并通过基于图的消息传递来相互更新它们的表示，作者们在 CommonsenseQA 和 OpenBookQA 数据集上评估 QA-GNN，并展示其相对于现有 LM 和 LM＋KG 模型的改进，以及其执行可解释和结构化推理（例如正确处理问题否定）的能力。

https://arxiv.org/pdf/2104.06378v1.pdf

683，基于通用模式理论的卷积神经网络可解释性

Erico Tjoa，Guan Cuntai

南洋理工大学，阿里巴巴

2021.2.5

已有的工作和研究为深度神经网络（DNN）的可解释性提供了许多见解和贡献，但现有理论仍然无法完全理解和解释 DNN。提高 DNN 的可解释性具有很多好处，例如设计可靠性更高的方法，以及更好地对算法进行维护和改进。由于数据集结构的复杂性会加大解决由 DNN 机制引起的可解释性问题的难度，因此本文提出使用一种由 Ulf Grenander 提出的模式理论，其中数据可作为基本对象配置，从而使

我们能够以组件方式研究 CNN 的可解释性。具体地，将扩展块附加到 Res Net 上来形成类似于 U-Net 的结构，从而使其可以在其 EB 输出广通道上执行与模式理论配置兼容的类似于语义分段的任务。通过这些 EB 模块来设计基于热图的可解释人工智能方法，以提取构成单个数据样本的单个生成点的解释，从而有可能减少数据集的复杂性对可解释问题的影响。包含上述模式理论元素的 MNIST 等效数据集可以让这种框架更加平滑，从而更加自然地通过图片生成的方式展示该理论。

684，脑机接口实践重大突破

国内首例接受生成 BCI 闭环响应刺激器植入的癫痫病患者顺利出院

浙大 BCI 张建民临床研究团队

在国内采用闭环神经刺激器被证明能有效控制临床应用中的癫痫发作后，张建民研究团队在脑机接口的应用方面取得了重要突破，他们于 2020 年 4 月 22 日在浙江大学医学院附属第二医院宣布：首例接受生一成 BCI 闭环响应刺激器植 x 约癫痫病患者顺利出院。

张建民教授说：这种闭环神经刺激器是一项基于脑机接口（Brain-Computer Interface, BCI）的先进技术，在植入后刺激器能够有效监测到癫痫发作信号，并对神经提供具有治疗效果的电刺激。这种刺激器采用无线充电技术供能。

685，欧盟出台人工智能系列新规，强调可信、安全和创新

2021 年 4 月 1 日，欧盟委员会提出新规的规则和行动，旨在使欧洲成为全球可信人工智能中心。

欧盟规则具体包括：①人工智能法律框架，即制定统一的人工智能规则（人工智

能法）并修正某些联合立法行为，②新的协调计划，即 2021 年人工智能协调计划审查，③针对机器的新规，即针对机器产品的立法提案。

①②将确保欧盟企业和公民的安全和基本权利，同时加强欧盟对人工智能的使用、投资和创新；针对机器的新规则将起到补充作用，通过调整安全规则提高用户对新一代多功能产品的信任度。

1），法律框架，打造可信人工智能

①风险不可接受

②高风险

③风险有限

④风险最低

2），协调计划：实现卓越人工智能的方法

3），机器新规：重视机器产品安全和创新。

686，用于时序模式学习的神经拟态算法——硬件协同设计

Haowen Fang，Brady Taylor，Ziru Li，Hai Li，Zaidao Mei，Qinru Qiu

锡拉丘兹大学，2021.4.21

神经拟态计算和尖峰神经网络（SNN）模仿生物系统的行为，并以其潜在的高能效执行认知任务的特性而引起了人们的兴趣。但是，某些已经证实对信息处理至关重要的因素（例如时间动态性和尖峰定时）经常被现有的工作所忽略，从而限制了神经拟态计算的性能和应用。一方面由于缺乏有效的 SNN 训练算法，时序神经动态信息难以被利用起来。许多现有算法利用统计学知识处理神经元激活。另一方面，利用时序神经动态信息也对硬件设计提出了挑战。突触表现出时间动态，

用作保存历史信息的存储单元，但通常会简化为与重量的关系。当前大多数模型将突触激活整合到某些存储介质中以表示膜电位，并在神经元发出尖峰后建立膜电位的硬复位。这样做是为了简化硬件，仅需要“清除”信号即可擦除存储介质，但会破坏存储在神经元中的时间信息。本文导出了一个针对泄漏整合和火灾神经元的有效训练算法，该算法能够训练 SNN 来学习复杂的空间时间模式。同时本文的算法在两个复杂的数据集上获得了极具竞争力的准确性。此外，本文还万达过新颖的时间模式关联任务证明了文中模型的优势。使用该算法进行代码签名，为基于忆阻器的神经元和突触网络开发了 CMOS 电路实现，该电路保留了关键的神经动力学，并降低了复杂性。对神经元模型的这种电路实现进行了仿真，以证明其对具有自适应阈值的时间尖峰模式做出反应的能力。

687，脉冲神经网络中无监督模式识别的权重发散促进原理

Oleg Nikitin，Alex Kunin，OlgaL ukyanova

俄罗斯科学院，2021.4

信号处理任务与生物神经元之间的并行性加深了人们对输入信号识别自组织优化原理的理解。本文讨论了生物学和技术系统之间的相似之处，提出了对著名的 STDP 突触可塑性规则的补充，以将权重调整指向与背景噪声和相关信号之间的最大差异相关的状态。物理上限制权重增长的原理被用作这种控制权重改变的基础。有人提出，生物突触的直接修饰受到可塑性发展所需的生物化学“物质”的存在和生产的限制。在本文中，有关信噪比的信息用于控制此类物质的产生和存储，并驱动神经元的突触压向具有最佳信噪比的状态。本文进行了使用不同输入信号体制的几个实验，以了解所提出方法的功能。

688，识别、调整和集成：将知识图谱与常识推理任务进行匹配

Lisa Bauer，Mohit Bansal

北卡罗来纳大学教堂山分校

2021.04.20

https：//arxiv.org/pdf/2104.10193v1.pdf

在解决常识性推理任务中的知识空白方面，通过将外部知识整合到常识性推理任务中可以较为有效的解决这些任务中的一些（但不是全部）问题。为了使知识集成达到最佳性能，选择与给定任务目标完全吻合的知识图谱（KG）至关重要。

作者们提出一种方法来评估候选 KG 可以正确识别并准确填补任务推理的差距的能力,他们将其称为KG与任务的匹配。作者们分3个阶段显示KG-to-task匹配：知识-任务识别、知识-任务对齐和知识-任务集成。

作者们还通过常识性探针分析了基于变压器的KG-to-task模型，以测量KG集成前后在这些模型中捕获了多少知识。从经验上讲,作者们调查了 SocialIQA(SIQA)，Physical IQA（PIQA）和MCScript2.0数据集的KG匹配项，其中3个各种KG：ATOMIC，ConceptNet，以及基于WikiHow的自动构建的教学KG。通过他们提出的方法，他们能够证明针对事件推理的KG ATOMIC是SIQA和MCScript2.0的最佳匹配，并且他们通过人工评估验证了概念类ConceptNet和WikiHowbased KG在所有3个分析阶段均是PIQA的最佳匹配。

689，日本大阪大学、英国利物浦大学的科研人员通过机器学习算法研发新材料

日本大阪大学教授利用 1200 种光伏电池材料作为训练数据库，通过机器学习算

法研究光伏电池高分子材料结构和光电感应之间关系，成功在 1 分钟内筛选出有潜在应用价值的化合物结构（传统方法需要 5～6 年）；英国利物浦大学科研人员研发一款机器人，自主设计化学反应路线，8 天内完成 688 个实验，找到一种高效催化剂以提高聚合物的光催化性能。

690，半监督文本分类中的虚拟对抗性训练增强注意机制的鲁棒性和可解释性

Shunsuke Kitada，Hitoshi Iyatomi，2021.4.18 发表

日本 Graduate 科学与工程学校，Hosei 大学

本文提出了一种基于虚拟对抗训练（VAT）的注意力机制的新通用训练技术。VAT 可以在半监督的情况下从未标记的数据中计算出对抗性扰动，以用于先前研究中已报告的易受扰动的注意力机制。经验实验表明，本文的技术①与基于对抗训练的传统技术以及基于 VAT 的技术在半监督环境下相比，提供了明显更好的预测性能；②证明了与单词重要性相关性更强，并且更好与人类提供的证据一致；③随着无标签数据量的增加，性能有所提高。

691，3D 脑肿瘤医学图像分割网络中的视觉可解释性

Hira Saleem、Ahmad Raza Shahid、Basit Raza

2021.4.26

巴基斯坦国家人工智能中心

医学图像分割是一项复杂而又重要的任务，它是医学诊断中最重要的环节之一。基于 3D 卷积神经网络（3DCNN）的模型在脑肿瘤图像分割方面取得了显著成果。然而由于神经网络的黑盒子性质，很难解释模型给出预测结果背后的基本原理，

因此在医疗健康领域，集成此类模型以做出有关诊断和治疗决策的风险很高。因此在医学领域部署深度学习模型时，需要准确且透明的预测。在本文中，我们通过一种扩展的解释性技术来生成 3D 视觉解释，来提高 3D 脑肿瘤分割模型的可解释性。我们首先分析了无梯度可解释性方法相较于梯度依赖方法的优势。然后我们解释了分割模型对输入磁共振成像（MRI）图像的操作，并研究了该模型的预测策略。

我们还评估了其他多种针对医学图像分割任务的可解释性方法。为了证明我们的视觉不包含多余的噪声虚假信息，我们定量地对扩展方法进行了验证测试。该模型捕获的信息与人类专家的领域知识是一致的，从而使其风险性更低。最后我们使用 BraTS-2018 数据集训练 3D 恼肿瘤分割网络并通过可解释性实验以生成预测结果的视觉解释。

692，基于近似误差反向传播的脉冲神经网络

Matteo Cartiglia、Germain Haessig、Giacomo Indiveri

瑞士苏黎世大学神经信息学研究所，苏黎世理工学院

2021.4.23 发表

脉冲神经网络（SNN）在低功耗传感处理和边缘计算硬件平台的设计方面展示了广阔的前景。然而，在这样的架构上实现片上学习算法仍然是一个开放性的挑战，特别是对依赖于反向传播算法的多层神经网络而言。在本文中，我们提出了一种基于 SNN 的学习方法，该方法使用局部权重更新机制进行近似的误差反向传播操作，可以与混合信号模拟 / 数字神经形态电路兼容。我们设计了一种网络结构，该体系结构使突触权重更新机制能对误差进行跨层的反向传播，并提出了一种网

络，该网络可以通过训练来区分具有相同平均放电率的两种模式。这项工作是迈向具有片上学习能力电路的超低功率混合信号神经形态处理系统设计的第一步，该电路可以通过训练来对不同的时空模式进行识别（例如基于事件的视觉或听觉产生的时空模式）。

693，中美技术对抗前沿：中国 BCI 技术再次取得突破

Li Xuanmin，Qi Xijia ，2021.4.27

天津大学脑科学中心，浙江大学医学院

脑机接口（BCI）是个重要的生物科学领域，其在工业领域具有重要应用价值，具有数万亿美元的市场，也是中美技术竞争对抗的前沿领域。目前中国研究团队正在自主开发用于 BCI 的芯片。天津大学神经工程团队正在研发第二代“Brain Talker”芯片具有更低功耗，并可提供更高的片上系统集成度。通过佩戴覆盖有敏感电极的大脑电极帽，并插入芯片，一个人可以用自己的思想打字。通过对佩戴者的大脑信号进行解释，即可在屏幕上显示相应的文字，而无需通过键盘输入。天津脑科学中心副主任徐敏鹏称，研究团队使用该芯片可以从脑电波信号中捕获高质量的大脑意图信息，从而满足应用需求。

第一代“Brain Talker”芯片已于 2019 年发布，然而目前该技术离达到商用还有一定距离。天津大学团队是中国加快突破 BCI 技术瓶颈的典型例子，因为目前中国在芯片和处理器等核心技术方面仍然依赖要从美西方进口，许多 BCI 产品仍然需要依赖于进口芯片和材料的支持来改进产品的体积和性能。

在微创 BCI 领域，中国与外国竞争对手还存在差距，以特斯拉 CEO 埃隆-马斯克创立的Neuralink 为代表的美国公司更胜一筹，而中国的技术障碍存在许多方面：

高灵敏度传感器、高保真神经元信号收集器、高精度微环境控制器等。

在无创 BCI 领域，中国取得的一些技术成果也是世界一流的。去年，天津大学团队使用混合编码方法创建了具有世界上最大命令集的高速 BCI 系统，该系统能够处理 108 条计算机指令，大约是其他脑机系统的 3～4 倍。复旦大学展示了其自制的首款用于动物的远程 BCI 芯片，该芯片的重量仅为国外同类芯片的一半。天津团队还与中国宇航员培训中心合作，探讨 BCI 技术与智能机器人联合进行航空航天探索。

目前，由浙江大学开发的闭环神经刺激器在治疗癫痫病方面取得重大成果。闭环神经刺激器是一项基于 BCI 的先进技术，可在早期识别癫痫发作，并且能够响应癫痫发作而提供治疗性电刺激。该技术打破了美国对于闭环神经刺激器的技术封锁。此外，该产品相较于国外同类产品具有更小的体积和重量，通过无线充电技术供能，具有更长的使用寿命。

694，比较类脑计算和传统计算的不同运行方式

	类脑计算运行方式	传统计算运行方式
信息源	神经电脉冲信号和化学信号	数字信号（或模拟信号）
编码方式	采用稀疏脉冲时序编码机制	由数字源代码变换为 0，1 机器码进行编码
计算模式	计算（神经元）和存储（突触）是一体化的，融合一起 存在三维泛连通性，基于脉冲的事件驱动型的随机计算	计算(处理单元)和存储单元是分离的，无法模拟三维连通，受限于二维连接，构建确定性计算
传递方式	通过由仿脑自然神经元＋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	传统网络权重连接＋激活方式，对机器码进行信息传递
基于不同特征的计算方式	类脑计算（拟态计算）系统的运行方式，打破冯-诺伊曼计算架构	传统计算机的运行方式符合冯-诺伊曼硬件和软件计算架构 而且一、类脑算法，低能耗，二、传统算法，能耗大。

695，关于通用人工智能的讨论

在人工智能国内外跟帖讨论中，钟义信、张钹等均谈到人工智能发展前沿为通用人工智能，但至今尚未突破具体解决方案。李德毅院士提出通用人工智能“十问”，他提问如何理解通用人工智能？王迪兴曾答复“十问”，他在谈到通用人工智能时，也仅停留在议论的层面上，并未涉及具体解决方案！但这里的问答，也可供探索者参考！

问：如何理解通用人工智能？我们应该不应该把通用智能理解为“全知全能”或者单项超强智能？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 。遗憾的是发展六十多年的人工智能没有能夠更靠近人的原始智能。

答：通用人工智能需要全新的理论基础，从哲学层面讲，需要本体论、方法论、认识论的统一，从自然科学的角度讲，需要系统论、信息论、控制论的统一。从计算理论角度讲，巴贝奇与图灵计算模式均不适用，需要奠定新的计算理论基础，统一时间与空间计算理论，且与脑科学统一。

根本性障碍在于目前数学描述方法对于结构化的多因果互为因果关系描述无能为力，需要创建一门新的数学——结构数学，其理论基础是准全息系统论，其定量描述是全息结构计算模型。

基于它可设计 2-16-256 进制的类恼计算机（我们已设计实现 16 进制计算机），不仅计算速度能达到指数级计算能力增长，关键是性能更符合人脑功能机制。

智能的最本质特性，是子系统或功能模块基于多因多果、互为因果关系建立实时、共时交互作用的关系模式。智能是其若干功能模块互补交互作用的结果。相对于传统计算机，其功能特征是读、写、算同步、储算一体化、址与数据统一。

外设之间能够多对多并行实时双向同步交互作用，体现内在的统一性，即运算、交换、控制、双向立交总线功能统一一体化。大脑和遍布人体的神经网络本来是一体的东西，但在现实世界，计算机和网络却人为的被分成两个相对独立的东西，丧失了内在统一性。智能最本质的东西，是记忆单元与各功能单元具有内在自组织作用关系。

不仅存储单元之间具有确定性的互为因果作用关系，且能形成紧密交互作用的更大规模的超循环！所谓超循环是每一个作用因子的输出是另一个作用因子的输入，这是智能、生命产生的基本前提。

模拟人脑智能，即模拟类脑神经元网络的结构及功能，多因多果、互为因果作用关系模式，是生命及智能的基础。

至于通用人工智能的目的不应该是构造一个与人脑完全一样的东西，完全相同的克隆人就行了！也并非全知全能，人也非全知全能！说到底我们不是构造具有独立意志的智能异己！只有不同才会超越！

696，人工智能如何在非冯架构上表现？

问：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？

答：存在非冯计算结构是肯定的，基于准全息系统论与结构计算模型，我们已设计实现 16 进制类脑计算机，在此基础上设计 256 进制计算机也完全没有问题！这是典型的非冯结构计算机。不仅能体现储算一体化特征，更能体现环境和背景知识双驱动的本质特征。

人工智能迄今已产生若干理论和技术流派，但用硬件实现的成果较少，显现智能的突破性进展有限。迄今为止，按生物神经网络（BNN）巨量并行分布方式构造的各种人工神经网络，并未体现人们所期盼的智能。专家系统在一个价段的成应用后逐步显露其局限性。历时 10 年耗资 540 亿日元的第五代计算机，在研制可实用化的计算机方面并没有完成预期目标。在涉及现场问题处理及实时获取知识方面遇到瓶颈，过多的规则和预先规定限制了系统的柔性和适应性，如模式识别能力较差，增强处理问题的适应性成了首要难题。在事实驱动和目标引导的双向推理方面未能引入实时交互作用机制。神经网络、知识工程、行为机制和现场学派都是从不同侧面模拟局部人脑智能，不能从系统构成的理论层面提供有益启示。人脑智能模拟必须基于自然系统理论才能解决。软件代替不了符合自然原理的硬件。寻求自然计算原理的突破是构造类脑计算机的唯一正确选择，试图用软件模拟人脑智能此路不通！智能是多层次、多子系统的功能互补效应，不能在一个功能层次完整体现。

设计类脑计算机既解决计算机与背景信息相融统一问题，又解决计算机与人脑近似同构同功问题。

类脑计算或智能模拟必须解决三大问题：一，解决状态之间的自组织模式（存储

单元的内在联系）问题，二，解决计算原理及效率（与人脑同构同功）问题，三，解决多功能层次的统一运行机制及双向交互作用（功能耦合）问题。同时解决计算机与环境及背景信息分离的问题，解决传感、效应子系统与核心信息处理机制分离问题，解决储算分离问题，解决总线瓶颈、二值逻辑、确定性及形式化悖论等问题，这些问题绝非传统理论及技术能够解决的。软件本质上是发挥硬件功能，硬件不具备的功能软件同样无能为力，因而不能指望软件模拟人脑智能。

计算机存储单元之间没有内在联系，子系统功能耦合与状态的自组织机制分离，状态转换与环境作用机制分离。而人类状态关系结构与状态记忆、处理结构是统一的。在转换状态时，状态与环境及背景信息统一，是智能本体（自然）状态关系的自组织。基于这种自组织，状态转换的各层次并非互相分离，而是实时透明传递及处理信息的。

搞人工智能或类脑智能有一个目的性问题，即我们不是构造完全相同的人脑，而是仿脑。两者的本质区别在于人脑有历史积淀，已经形成相对的固定模式，而仿脑可以跳出这种相对固定模式，完全一样反而不能在某一方面超越，仅仅是一样而已，类恼只有超越才能体现更大价值。

模拟人脑智能，只要与人脑具有逻辑同构性，就可以在记忆、逻辑推理、知识的条理化方面超越人脑，因为人脑要受环境及教育程度的限制，而仿恼没有这样的限制，它可以集中全人类的智慧于一体！因而超过单一个体人脑是理所当然的事！仿脑最重要的一点就是具有超强的记忆能力，基于逻辑内涵与外延的逻辑一致性，仿脑可以把人类历史上、及所有人的知识集中在一起，且比人脑的信息处理速度更快、更精确、更具系统性！因而比人处理信息的能力更强！

697，商汤科技研发的人脸识别在全球领先

德国《商报》在 2021.4.3.一篇“中国对全球经济攻势”的文章中谈到商汤科技（ Sensetime）的人脸识别一体机，0.3 秒极速验证准确率 99.99%。

2 月下旬，谷歌前 CEO 埃里克-施密特受拜顿总统委托，在评估中美人工智能发展水平时也谈到：“在人脸识别上，中国超过了美国，在全球是顶尖的”。

698，点评中国数字人民币试点

2019 年底数字人民币试点、测试相继在深圳、苏州、雄安、成都四地及北京冬奥会场启动，到 2020 年 10 月，增加了上海、海南、长沙、西安、青岛、大连六个试点测试地区，即中国数字人民币试点有序扩大至“10+19”。试点工作以受邀白名单小额交易为主(目前参与人数、参与筆数、净兑换的金额，总体上还较小，试点场景现在覆盖生活缴费、步饮服务、交通出行、购物消费、政务服务等多个领域，工、农、中、建、交、邮储等中国六大国有银行向公众申请白名单，京东自营线上已支持数字人民币购物，。今后将进一步探索数字人民币应用模式，强化数字人民币通用性和普惠性，完善产品功能和应用性，提高系统的安全性、稳定性，并改善跨国转账系统。

全球数字货币研发加速，日本、法国、俄罗斯等国正在加紧研发。美国政府关注包括数字人民币在内的全球数字货币研发落地，担心当前全球金融体系结构面临的直接挑战。

699，第五代移动通信 5G

2020 年，全世界建设约 100 万个 5G 基站，中国占比约 70%。

2020 年中固已经建成 72 万个 5G 基站。

5G 不仅在速度上提升，同时把低时延、低功耗、高可靠的愿景引入移动通信中，为工业互联网、自动驾驶和无人驾驶、智慧城市、有关商业模式的需求提供了可能。

天地一体化，进行广域全覆盖，应是 6G 发展的使命和重要愿景。6G 将促使互联网底层架构的改造，面向智能互联网。

700，2021 中国开源发展蓝皮书发布

蓝皮书全面阐述 2021 年中国开源总论、发展、开发者，开源社区和开发机制，开源企业和商业模式，开源组织和自治激励，开源发展面临的机遇、风险和挑战，国人对国际开源资源的贡献，开源人士和组织对技术、经济、社会、文化的贡献，开源的发展预期，以及开源教育、立法、知识产权保护、标准化、开源基金会、风险投资等各方面建设，在已取得重大成绩的基础上要更上一层楼。

蓝皮书收录了国际开源大师对中国开源发展的点评：中国开源发展很快，如今已接近或达到世界先进水平，一些企业开始进入世界领跑者行列，还涌现出杰出的开源领袖;：许多企业以亲身体验证实：当今开源已成为全球的一种创新和协同模式；许多研发者以亲身体验证实：当今开源已成为支撑深度信息技术发展的基

础。

蓝皮书谈到党和政府十分重视开源，开源已被写入十四五规划纲要：“支持数字技术开源社区等创新联合体发展，完善开源知识产权和法律体系，鼓励企业开放软件源代码、硬件设计和应用服务”。

701，迈向可解释和可转移的语音情感识别：基于潜在表示的特征、方法和语料库分析

Sneha Das、Nicole Nadine Lønfeldt、Anne Katrine Pagsberg、Line H. Clemmensen

丹麦科技大学，2021.5.5

近年来，语音情感识别（SER）已经在从医疗保健到商业的多个领域得到广泛应用。除了信号处理方法外，SER 的方法现在还使用深度学习技术。但是对语言、语料库和记录条件进行概括仍然是该领域的挑战。此外，由于深度学习算法黑盒子性质，模型和决策过程缺乏解释性和透明性成为了新的挑战。当将 SER 系统部署在影响人类生活的应用程序中时，这一点至关重要。在这项工作中，我们通过对所提出的 SER 系统的决策过程进行深入分析来解决这一问题。为此我们提出了基于不完全和去噪自动编码器的很低复杂度 SER，对于四类情感分类该编译器的平均分类精度达到 55％以上。在此之上，我们调查了潜在空间中的情感聚类，以了解语料库对模型行为的影响并获得对潜在嵌入的物理解释。最后，我们探讨了每个输入功能对 SER 性能的作用。

702，神经形态计算中的动态可靠性管理

Shihao Song、Jui Hanamshet、Adarsha Balaji、Anup Das、Jeffrey L. Krichmar、Nikil D. Dutt、Nagarajan Kandasamy、Francky Catthoor 2021.5.5

加州大学、德雷塞尔大学、比利时微电子研究中心

神经形态计算系统使用非易失性存储器（NVM）来实现高密度和低能耗的突触存储。操作 NVM 所需的升高的电压和电流会导致每个神经元和硬件中的突触电路的 CMOS 晶体管老化，从而使晶体管的参数偏离其标称值。激进的设备扩展会增加功率密度和温度，从而加速老化，对神经形态系统的可靠运行造一成挑战。现在的以可靠性为导向的技术（假设最坏的工作条件）以固定的时间间隔周期性解除硬件中所有神经元和突触电路的工作负载，而实际上并未在运行时跟踪它们的老化情况。为了减轻这些电路的压力，必须中断正常操作，这会在脉冲产生和传播中引入等待时间，从而影响脉冲间隔和性能，例如精度。通过设计智能运行时管理器（NCRTM），我们提出了一种新的体系结构技术，以缓解神经形态系统中与衰老相关的可靠性问题，该管理器可动态响应神经元和突触电路在 CMOS 晶体管中的短期衰老，从而缓解神经元和突触电路的压力。该管理器在执行机器学习工作负载期间，以满足可靠性为目标。NCRTM 仅在绝对必要时才对这些电路施加压力，否则会通过在关键路径之外安排压力操作来降低性能影响。我们使用神经形态硬件上的最新机器学习工作负载来评估 NCRTM。结果表明，NCRTM 显著提高了神经形态硬件的可靠性，对性能的影响很小。

703，SpikE：基于脉冲的多关系图数据嵌入

Dominik Dold、Josep Soler Garrido 2021.4.27

西门子

尽管最近成功地将基于脉冲的编码与错误反向传播算法相协调，但脉冲神经网络仍主要应用于感官处理产生的任务，这些任务在视觉或听觉数据等传统数据结构

上运行。可以在工业和研究领域得到广泛应用的丰富数据表示形式就是所谓知识图谱--一种基于图的结构，其中实体被描述为节点，而它们之间的关系被描述为边缘。可以使用知识图谱嵌入算法在这些信息密集的环境中进行上下文感知的预测。我们提出了一种基于脉冲的算法，其中图中的节点由神经元种群的单个脉冲时间和种群之间的峰值时间差的关系表示。学习这样的基于脉冲的嵌入仅需要有关脉冲时间和脉冲时间差的知识，就可以与最近提出的用于训练脉冲神经网络的框架兼容。提出的模型可以轻松地映射到当前的神经形态硬件系统，从而将对知识图谱的推理移入这些体系结构的领域，从而为该技术打开了一个有前途的工业应用领域。

704，老年人和脑机接口：一项探索性研究

日本 PolishIT 研究院

2021.4.5

这项探索性研究是针对老年人的智能家居技术（SHT）背景下研究无创脑机接口的可能性。

NOTE:

NOTE: