

Matrix

NO.67

2023年10-12月

25周年特刊

“智启未来”

系列署名文章

“智汇对话”

我今年25岁，

与微软亚洲研究院同龄

01 智启未来

刷新自我，引领计算的新范式	2
创新智能环境带来无线通信与感知的新视角	5
媒体基础：打开多模态大模型的新思路	8
跨学科合作构建具有社会责任的人工智能	11
人工智能基础创新的第二增长曲线	14
当 AI 遇见大脑：电脑与人脑协同“进化”	18

02 独特记忆

25 个关键词认识 25 岁的微软亚洲研究院	22
------------------------	----

03 智汇对话

周礼栋对话比尔·盖茨：深入的科学研究比以往任何时候都更加重要	25
周礼栋对话青海大学校长史元春：跨学科创新，与 AI 共成长	27
周礼栋对话 Chris Bishop：转变思维方式，重塑科研文化	29
Ashley Llorens 对话薛澜：推动人工智能正向发展，跨领域与国际合作至关重要	31
洪小文对话图灵奖获得者 Raj Reddy：人工智能将改变我们学习方式	34
AI 与社会、科学的协同：跨学科研究的潜力与挑战	36

04 25 周年系列庆祝活动精彩瞬间

39

05 我今年 25 岁，与微软亚洲研究院同龄

41

"智启未来"

系列署名文章

2023 年是微软亚洲研究院建院 25 周年。25 年来，微软亚洲研究院探索并实践了一种独特且有效的企业研究院的新模式，并以此为基础产出了诸多对微软公司和全球社会都有积极影响的创新成果。一直以来，微软亚洲研究院致力于创造具有突破性的技术。在人工智能时代，微软亚洲研究院将为计算新范式奠定基础，并为人工智能和人类发展创造更美好的未来。

在“智启未来”系列署名文章中，微软亚洲研究院不同研究领域的领军人物分享了他们对人工智能、计算机及其交叉学科领域的观点洞察及前沿展望。希望此举能为关注相关研究的同仁提供有价值的启发，激发新的智慧与灵感，推动行业发展。

刷新自我，引领计算的新范式

"赋予微软亚洲研究院的新使命是致力于通过跨学科和跨界研究，创造具有重大影响的突破性技术，为未来几十年的计算新范式奠定基础，并为人工智能和人类发展创造更美好的未来。"

——周礼栋，微软全球资深副总裁、微软亚洲研究院院长



微软亚洲研究院院长周礼栋

在 2023 年即将远去的这一刻，我不禁会产生这样一种遐想：若干年后，未来的人们会回想起这一年中大家兴奋地学习使用大语言模型 (LLMs) 和各种 Copilot 的情景，并感慨道：“改变就是从那时开始发生的”。尽管 2023 年的我们对人工智能的了解还在初级阶段，但已确信它将与人类的命运紧密相连。

对微软亚洲研究院而言，2023 年则更多了一些新的里程碑意义：我们迎来了建院 25 周年。回首过去的 25 年，我们可以自信且自豪地说，微软亚洲研究院探索出了一套企业研究院的有效模式，孕育了诸多对微软公司和全球社会都具有重大影响创新成果，见证并推动了亚洲乃至全球科技的发展，取得了足以让我们每位伙伴骄傲的成就。

在最近与研究院同事们的交流中，我们不约而同地达成了一个共识：与其“追忆往昔”，我们不妨把目光投向下一个 25 年，以展望未来的姿态来致敬这次意义非凡的周年庆。因此，近来我们一直在思考：在面对新的技术发展浪潮时，过去那些使我们成功的思维、文化与工作方式，在坚守与传承的同时，是否应该做出一些突破性的改变？

答案是肯定的。我们希望不断刷新自我，继续创造具有深远影响的新成果，为人类社会的福祉做出贡献。

刷新目标使命，定义微软亚洲研究院下个 25 年

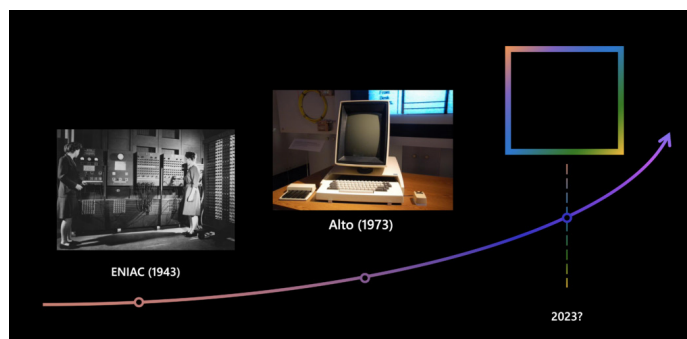
无论是作为微软全球研发体系的一员，还是作为企业研究院的代表，微软亚洲研究院都具备一些与众不同，甚至独一无二的特质。背靠微软公司这一全球化的平台，又身处信息产业蓬勃发展的亚洲地区，微软亚洲研究院得以融合东西方创新文化的精髓。自建院以来，我们就秉持开放、自由、多元与包容的文化精神，吸引全球优秀科研人才汇聚于此。这也是微软亚洲研究院能够在 25 年中始终站在科技浪潮前沿的原因。

每位微软亚洲研究院的伙伴都希望我们能在未来几十年里继续保持我们的优势和独特性。而这正是微软亚洲研究院寻求刷新自我的出发点：我们已经身处一个不同的时代和环境。就像那些能够保持长期竞争力的组织一样，我们必须与时俱进，做出必要的改变。这不仅仅包括增加新的研究项目，或是简单地改变某些流程，更关键的是，我们要着眼于下一个 25 年的长远视角，思考微软亚洲研究院应该向什么方向前进，并确保我们的步伐始终与我们的愿景相一致。

至于下一个 25 年，我认为变化、趋势、目标……都将紧紧围绕计算新范式展开。

计算机领域的范式转换周期大约是三十到五十年，每一次范式变革都会重新定义计算机在未来数十年里如何影响和改变人类社会。上一次范式转换的起点正好在五十年前，施乐帕克研究中心（Xerox PARC）推出了 Alto 计算机，开启了从工业计算向个人计算的范式转换，并由此激发了众多彪炳史册的创新。

值得一提的是，在 Alto 诞生的时候，人们并没有在第一时间将其定义为个人计算时代的开始。但是包括 Xerox PARC 在内的研究者都意识到需要做些不同以往的事情。后来，随着图形界面、以太网、鼠标、激光打印机等创新技术的出现，大家才逐渐看清这一轮新范式的轮廓。



现在，世界又一次来到孕育新一代计算范式的关键节点。新的计算范式将赋能人类生活和工作的方方面面，给各行各业带来颠覆性的变革，也将催生众多新的机遇。我认为，此时最有意义的工作之一，莫过于思考我们应该以怎样的方式进行研究，从而有助于定义新范式——这不仅涉及到研究和创新，更关乎人类社会的未来。

我们认为，赋予微软亚洲研究院的新使命是致力于通过跨学科和跨界研究，创造具有重大影响的突破性技术，为未来几十年的计算新范式奠定基础，并为人工智能和人类发展创造更美好的未来。

刷新研究方式，确保向目标迈进

我们对研究方式也进行了重新思考，以帮助我们更好地达成使命。事实上，微软亚洲研究院自成立以来一直不断调整我们的思维方式与创新模式，未来我们也将做出更多变革。

一、积极投入跨学科和跨领域研究

在微软亚洲研究院看来，未来具有重大意义的突破性成果，将来自跨领域的交叉融合。我们已经与多个自然科学领域展开合作，而为了让技术服务于解决人类社会问题，与社会科学学科的合作也将是必不可少的。

我们的目标不仅仅是打破计算机科学与其他学科间的壁垒——即使在计算机科学领域内，也应该摒弃以往那种只专注于自己的细分领域的做法。大语言模型就是一个很好的例证：孤立的领域将不复存在，更强的机器智能将来自跨学科与跨领域的深度融合。

二、连接全球的卓越人才和智慧

在新的范式下，当我们提到“全球人才”，其范围将不再局限于“计算机人才”。随着跨领域的持续深入，我们需要与各学科和行业的人才建立更紧密的联系，并为不同领域的人才搭建有效沟通的桥梁，才能跨越学科的边界，推动学科交叉与研究融合。

未来，微软亚洲研究院将继续打造更加广阔的国际化研究平台，与全球各地的优秀研究者、机构和大学建立更紧密的合作关系，推动知识的自由流动和技术成果的共享，通过汇集全球智慧，共同探索 and 解决世界面临的重大挑战。

三、创造具有重大影响的技术突破，着眼于具有社会意义的挑战

研发成果被应用于企业产品，甚至引领产品未来的发展方向，当然是了不起的成绩。但如果我们止步于此，将失去作为基础研究机构的独特性。前不久，微软联合创始人比尔盖茨先生到访微软亚洲研究院，我有幸主持了一场同他的对话。盖茨回顾了当初成立微软研究院的初衷，就是可以为人类的普遍知识做出贡献。

因此，微软亚洲研究院将致力于创造具有广泛社会影响的突破性技术。一方面，我们要确保技术具有足够的前瞻性，经得起时间的检验；另一方面，依托这些重大技术突破，我们能够在解决那些具有广泛和长远意义的技术、产业和社会问题上发挥不可替代的作用。

四、定义并引领计算的新范式

新一代计算范式是计算机科学研究，乃至更广泛的行业 and 人类社会共同面临的机遇。面对这样的机遇和挑战，我们必须行动起来，思考我们日常的研究和工作是否有助于深化对新范式的理解和定义，而不仅仅是专注于发表一篇论文，优化提升某项技术，

或是服务于某个产品。

定义和奠定新范式的基础将成为我们具有最大共识的战略方向。微软亚洲研究院希望能够在未来帮助各领域和行业建立起新范式的基本理论和底层逻辑，从而让所有人能充分把握这些新的机遇。

传承与刷新独特文化，激发每个人的潜力

微软亚洲研究院拥有独特且让我们引以为傲的文化。但我想强调的是，文化本身从来不是我们的目的，企业文化的真正价值在于激发个体的智慧与行动力。作为一个组织，微软亚洲研究院最关心的是，我们能否共同创造一个让每个人都能快速成长，并充分发挥潜力，产生深远影响力的环境。

基于组织的使命和对人才成长的承诺，我们致力于营造一种有利于激发灵感和创新力的研究文化：我们倡导对技术进步怀有远大抱负并富有极客创新精神，鼓励研究人员拓展研究的深度与广度，跨越领域界限；我们致力于打造利于培养长远愿景、倡导开放协作精神和多元包容思维的研究氛围；我们也不断探索将好奇心驱动与精准聚焦相结合的科研新范式。

显然，我们所追求的这种思维方式和文化氛围都有着一个共同的前提——消除任何不利于集体智慧和创新的障碍，让想法自由流动（flow），让最好的想法胜出。我们希望研究人员能在这样一个鼓励个性表达且被充分信任的环境中，迸发出更深远的灵感与思考。正如两度获得诺贝尔奖的莱纳斯·卡尔·鲍林（Linus Carl Pauling）所说：“获得好想法的最佳方式是先获得很多想法。”



我们正从多个方面入手来创造更加畅通无阻的沟通环境。比如，我们将尽可能避免自上而下的“要求式”对话，取而代之的是让研究员乐于表达和交流想法的形式。当他们感到自己是在主动参与而非被要求时，将更有动力和意愿表达新的想法。

其次，我们鼓励每个人积极发现未来的潜在机会并勇于尝试，即使那些机会听起来遥不可及，但这比展示已有成果更加重要。

正如我们无法用现有的知识来定义新范式，当我们勇敢地追求“未知”，才有可能在未来将其变为新的“已知”。

此外，我们为新员工建立“Aspire”社团，帮助他们快速适应科研工作和文化，并且通过各种自发的项目和活动快速成长。与此同时，新员工也为我们的文化注入了源源不断的活力。

聚焦引领新范式的研究领域

尽管各界对新范式的探索尚处在起步阶段，但微软亚洲研究院已经开始在一些具有潜力的研究领域展开探索，其中很多都是跨学科的合作。这些跨学科项目不仅对相关领域具有深远的影响，同时也提升了我们对于 AI 能力的理解和应用，有助于我们揭示新范式的奥秘。

基础模型迭代：在现有的“Transformer 基础网络架构 + Next Token Prediction 学习范式”这一基础架构之下，人工智能的发展主要依赖于增加训练数据量和扩大模型规模。然而，这种增长的天花板逐渐变得明显。我们设计的新的网络基础架构 RetNet，打破了 Transformer 无法实现的“低成本推理、高效率、同步训练”不可能三角，可以以更低的能耗及成本，实现人工智能效率和性能的指数级提升。

AI 与系统结合：AI 诞生于大规模高性能计算机系统之上。现在，我们正在研究如何利用 AI 助力计算机系统的创新和发展。在我们看来，AI 新的能力能为传统系统问题提供新的角度，例如：更智能高效地优化复杂系统的性能；更快速和智能的问题诊断；以及更便捷的部署和管理。AI 与系统相结合，也能为新系统的设计带来不同的范式。从芯片设计、体系结构创新、编译优化到分布式系统设计，AI 可以成为系统研究者的智能助手，甚至承担大部分工作。

创新智能环境：无线技术将不仅仅是通信的工具，还将成为 AI 发展和应用的“数据桥梁”和“第六感官”，使 AI 以超出人类的感官来全面理解物理世界，获得更强大的综合建模能力。然而无线技术正面临着理论和实践的双重挑战，尤其是通信范围和感知能力的限制。为此我们提出了全新的研究方向——为无线信道构建可调整的智能环境。这一创新可以利用低成本且易于部署的超表面装置使无线信号传输至原先难以触达的区域，还可以提升无线感知的精度，使其走向更广泛的应用场景。

人工智能与脑科学：通过深度融合人工智能技术与脑电信号、基因、血液等大脑关键组成要素，我们正在设计一系列特有的领域模型，这不仅将支持科研人员和临床医疗专家更好地获取大脑的反馈信息，发现脑部异常以及深入理解脑部疾病机理，同时也有助于我们有效了解人脑的思维方式和认知过程，从而给人工智能神经网络和机器学习算法的设计带来更多启发。

人工智能与传统文化：我们与甲骨文领域的专家合作，开发

了甲骨文校重助手 Diviner，首次将自监督人工智能模型引入到甲骨文校重工作中，帮助专家从这项原先完全依靠人工经验、既费时又费力的工作中解放出来。这一进展向学术界展示了人工智能技术在保护和传承富有社会意义的文化遗产研究中的巨大价值。

微软亚洲研究院对人工智能社会责任的重视，不亚于对其基础研究、智能性能和应用的关注。随着人工智能对社会的影响日益加深，如何确保人工智能的使用既遵循社会道德，又真正服务于人类，是所有人工智能研究者都必须深入思考的问题。因此，我们启动了“社会责任人工智能（Societal AI）”的研究，并与心理学、社会学、法学等社会科学领域的专家合作，以期能为人工智能制定可行的价值观标准，确保人工智能的发展和使用总能被置于人类利益的框架之内。

此外，我们还在人与 AI 交互、媒体基础、AI 在行业 and 科研的创新应用等领域进行探索。

写在最后

一个被广泛认同的观点是，基于洞察和预判的自我变革是组织长期保持竞争力和持续成长的必要手段。因此，我并不想将微软亚洲研究院正在进行或将来可能的自我刷新定义为“颠覆”。即便相隔数年再次造访，研究院依然会让你感到亲切，并迅速感受

到自己身在何处。我们相信微软亚洲研究院的文化具有足以穿越周期的生命力，使我们能够敏锐地观察世界的微小变化，将智慧应用于对新范式的理解和创新，让技术之美真正造福于人类社会。

本文作者：

周礼栋博士现任微软公司全球资深副总裁、微软亚太研发集团首席科学家、微软亚洲研究院院长，全面负责微软亚洲研究院在中国及亚太地区的研究工作以及与学术界和产业界的合作。

周礼栋博士是系统研究领域首屈一指的专家，多年来一直专注于推动可靠、可信及可扩展的分布式系统的理论研究和实践探索。作为微软在设计和开发大规模分布式系统方面的重要技术带头人，周礼栋博士主持设计和开发的系统支持着微软从搜索引擎、大数据基础设施、云可靠性和可用性到 AI 基础设施的主要服务。

周礼栋博士是电气电子工程师学会会士（IEEE Fellow）和国际计算机学会会士（ACM Fellow）。他还曾是 ACM 计算机系统会刊（ACM Transactions on Computer Systems）、ACM 计算机存储会刊（ACM Transactions on Storage）、IEEE 计算机会刊（IEEE Transactions on Computers）的编委会成员，并担任 ACM 软件系统奖项评选委员会（ACM Software System Award Committee）主席，以及 ACM 操作系统原理大会（SOSP）指导委员会（Steering Committee）成员。

创新智能环境带来无线通信与感知的新视角

“在人工智能时代，无线技术将不仅仅是通信的工具，还将成为人工智能发展和应用的“数据桥梁”和“第六感官”。然而，无线技术也面临着理论和实践的双重挑战，尤其是通信范围和感知能力的限制。为了突破这一瓶颈，我们提出了一个全新的研究方向——为无线信道构建可调整的智能环境。这一创新将提升无线通信和感知的性能与范围，创造更广泛的智能应用场景，推动社会朝着更加智能化的未来迈进。”

——邱锂力，微软亚洲研究院副院长



微软亚洲研究院副院长邱锂力

清晨醒来，立即伸手去拿手机已经成了多数人的习惯动作。手机和各种无线通信终端如今正在“包办”我们与世界的沟通交互，无论身处何地，只要有无线信号，就能即刻掌握全世界的新闻趣事，享受线上音乐带来的“情绪按摩”，观看各种精彩的视频和直播来放松身心，维系与亲朋好友间的交流沟通……这一切，都有赖于无线技术为人與人、人与物之间架起的沟通桥梁。

现在，无线技术也正在成为人类双脚跨进人工智能（AI）时代的一把钥匙。我们都相信人工智能的发展将极大地影响人类社会的发展方向。而要释放其更大的潜能，不仅需要人工智能模型本身的发展，也离不开各种与之相关领域的创新。这其中，无线技术无疑是最为关键的领域之一。

在许多人工智能应用场景中，无线通信技术都发挥着数据桥梁的作用。IoT 设备或智能终端生成的数据，需要强大、稳定的无线通信将其传输到云端的计算资源，才能实现实时的处理和分析；当我们需要更丰富的数据来训练人工智能模型时，无线感知则能够揭示那些视觉、语音、文字之外的现象，为人工智能提供更多元的数据类型，如视野外，穿过障碍物的透视，像呼吸心跳那样的微小移动，以及材料检测等等。

为了突破无线技术快速发展后所遭遇的理论瓶颈，我和微软亚洲研究院的同事们正致力于一项全新的研究探索——为无线信道创造可调整的智能环境，并已获得极具实用价值的成果。相信这一课题以及我们对无线通信和感知的其他研究，将有助于更多从业者开阔思维，探索无线技术在人工智能及其他更多场景中的应用，使之成为引领科技和社会发展的又一具强劲引擎。

无线通信逼近理论极限，通信范围成最大挑战

无论我们多么期待更多人工智能应用的落地，都无法忽视强大且广泛覆盖的无线通信是实现这一目标必要且无可替代的前提。作为云端计算服务器和边缘设备之间不可或缺的数据纽带，学术界对其愿景自然是希望它总能让终端应用实时获得云端的分析和响应。遗憾的是，现有的无线通信技术还不足以满足这一愿景。其中，通信范围不足是我们亟待解决的首要挑战。

Shannon 公式告诉我们，通信速率取决于发射和接收设备以及它们之间的无线通道，速率随着发射和接收天线的数目和质量增加而提升。回顾过往的一次次里程碑式的技术创新，都是在这理论指导下实现的。例如 4G 时代的 MIMO（多发多收），5G 时代的毫米波及大规模 MIMO。我们也做了不少探索和创新，如用压缩感知来估计无线通道以支持大规模 MIMO 编解码，分布式 MIMO 系统，节能 MIMO，高速运动时无线通道预测等。

但也正如 Shannon 公式所描述的，高速率需要大带宽。高频率有更大的带宽，但频率越高也意味着信号衰减越快，通信覆盖范围缩小。最简单的应对方法自然是提高发送 / 接收基础设施

的密度，比如设置更多的基站。但这无疑会降低无线通信的经济性，增加其部署成本。这一问题在 4G 及以前的无线通信部署中尚不突出，但步入 5G 时代后，提升频率来获取带宽与通信范围缩小之间的矛盾愈发逼近“临界点”，甚至已经对普及 5G 的进度造成影响，更遑论未来以太赫兹为发展方向的 6G 通信。

显然，如果我们不能以创新的方法突破这一瓶颈，无线通信就将长期受制于通信范围和部署成本之间的零和博弈。这不仅会给人工智能的广泛应用带来技术上的阻碍，而且还会使我们那些弥合数字鸿沟的行动止步不前。

$$C = B \log_2 \left(1 + \frac{S}{N} \right)$$

C : 信道容量, B : 信道带宽, S/N : 信噪比

突破传统思维， 为无线信道构建可调整的智能环境

以往我们对无线通信的改进都立足于“收”“发”两端，但如果我们重新审视整个无线通信系统，会发现除了发射机（tx）和接收机（rx）之外，无线系统还受到无线信道的制约——这在 Shannon 公式中其实也有所体现。

按传统思维，我们总是将无线信道视为由当前环境给定，是无法改变的。但现在我们意识到，“环境”或许是突破无线通信瓶颈的关键。于是我们产生了这样一个对未来的设想：调整环境更好地支持无线通信和感知！

至于实现方式，我们的考虑是在发射机和接收机之间放置一些低成本、无需电源的结构，然后根据应用需求来改变无线信道，从而为无线信号创造可调整的智能环境。例如在适当的位置设置反射面将无线信号反射至原先无法覆盖的区域，或是在环境中放置透射面来改变信号的轨迹。但普通表面需要严格遵循传统的反射定律和折射定律，对环境的调控很有限。给定一个环境经常无法找到合适的反射面或折射面来覆盖盲区。

超表面是实现我们这一设想的关键。它可以不遵循传统的反射定律和折射定律，对信号有更灵活的调控能力，可以任意改变出射波的方向，产生负折射率和完美透镜等神奇的现象。超表面由很多亚波长尺寸的单元组成，这些微小结构通过电磁波与其相互作用，产生局部的电磁响应，从而影响电磁波的传播特性。我们通过设计单元结构以及它们的排列方式来实现对电磁波的精确控制。

我和研究院的同事们开发了一款新型被动超表面 [1]，它无

需功耗，成本低，支持高带宽，高透射和 360 度相位调制。最大的问题在于被动超表面一旦制作完成即定型，无法动态调整出射波。考虑到在使用场景中接收端或发射端经常会移动，无法动态调制是一个严重的缺陷。为了解决这个问题，我们进一步研究与实践，将被动超表面与小型相控阵天线相结合，使其在保持被动超表面其他优点的同时，具备动态调整能力。

这一创新可以用于很多场景，例如 Wi-Fi 或蜂窝网络。我们可以将其用于低轨道（LEO）卫星通信领域，以缩小数字鸿沟。你可能很难想象，在 2023 年全球仍有多达三分之一的人口因为地处偏远、难以负担通信设施成本等原因，至今都没有接入互联网。虽然 LEO 通信提高无线网络覆盖范围的可行性已被验证，但其地面站的建设成本门槛对于一些地区来说依然难以逾越。而现在，我们的技术可以用于制造低成本地面站。



将小型地面站收发天线系统部署在地面，在高楼模拟卫星收发。经过验证，微软亚洲研究院团队设计的超表面能提升信噪比 6.8-45 倍。

经过初步摸索，我们已经认识到了通过超表面研究改造智能环境给无线通信带来的新潜力。其实改变出射波方向只是超表面的众多“本领”之一，除了支持无线通信，它在无线感知领域也有妙用。

创造智能环境，实现更易用的无线感知

无线电信号不仅可以用于传输“现成”数据，还可以感知并生成全新的数据，即无线感知（Wireless Sensing）。无线感知和无线通信的物理架构层基本一致，信号处理也相通，它源于无线通信的发展，如今也在反哺无线通信。例如通过精准定位信号接收体的位置，就可以把信号能量集中传输，以扩大传播距离；再比如 MIMO 这种多进多出的天线技术，可通过感知对周围环境进行建模，指导 MIMO 参数优化，从而在既定环境中实现最高效率……对于前面提到的通信范围难题，无线感知同样有很大的用武之地，在未来 6G 的研发中，无线感知几乎是必不可少的要素。

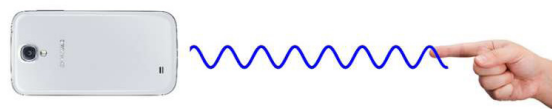
无线感知的作用远不止于支持通信。在我看来，无线感知有潜力成为改造人类生产、生活方式的颠覆性技术之一。试想，如果医生借助无线感知来获取病人的生理指标，那么病人就无需承受被插入各种监测设备的痛苦；又或者我们将无线感知技术应用用于野生动物保护，那么就可免去捕捉动物再植入传感器的麻烦。

在一些场景中，无线感知技术已被广泛应用。例如机场、车站及重要场合中必不可少的安检设备，医疗领域常用的 X-Ray（X 射线）和 MRI（核磁共振成像），都使用了无线成像技术。

无线感知极大延伸了人类感官，激发了我们对其未来应用的无尽想象。比如将其应用于智慧农业，了解农作物的生长状况，更有效地优化浇水、施肥和农药播撒来提高产量；再比如将其应用于食品安全和质量监测，无接触地检测食品是否出现变质、成分是否符合描述、添加剂是否合适等等。

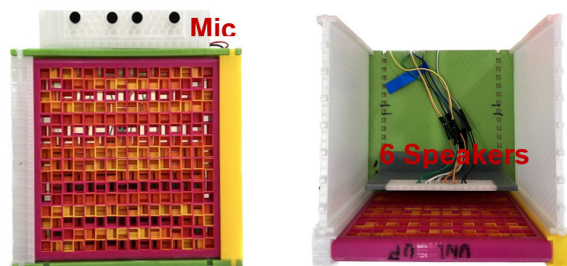
无线感知还能有力推动人工智能的发展和应用。它能为大模型提供远超人类感官范畴的数据类型，帮助人工智能用更多方式去学习和了解真实世界，以及为未来的机器人提供更强大的传感器，使之更精准地与环境互动。

近几年，对无线感知的研究和应用都在明显提速。一个显著的变化在于，过去无线感知需要使用非常昂贵的设备，而现在随着技术的进步，无线感知的普及门槛正在降低。有时候我们只需要使用普通的设备，就能实现高精度的追踪和成像。不久前，我们就开发了一种可以利用普通手机的内置扬声器和麦克风实现毫米级定位及声波成像的无线感知技术。相较于通过摄像头实现类似功能，声波不仅更加节能，能够更好地保护隐私，而且还具有一定穿透障碍物的能力，易用性和适用性都更好。



手机基于声波感知示意图

以往无线感知技术也是通过提升发射和接收设备的性能和天线数量来提高感知精度，会增加设备的成本、体积和能耗。既然无线感知与无线通信同源，那么构建智能环境的设想是否也有用武之地？答案是肯定的。首先，超表面有助于提高感知距离和精准度。比如我们设计的声音超表面 [2]，通过在普通扬声器和麦克风前放置一个低成本 3D 打印的声音超表面，就能大大提高感知距离和降低定位误差；其次，超表面还能把普通环境中感知不到的微小特征转变成更易感知的特征。这都有助于降低无线感知部署的难度，推动其走向广泛应用。



(a) 前视图 Front View

(b) 内部视图 View Inside

声音超表面

为无线通信和感知创造新的智能环境，为我们开启了新的探索维度和无限可能性。尽管我们尚处于起步阶段，任重而道远，但我们也期待与有志于此的伙伴们携手合作，共同迎接未知的挑战。我相信，将人工智能技术与应用场景的拓展相融合，有助于我们实现智能化、高效和可靠的无线通信和感知系统，从而推动诸如自动驾驶、智能医疗、工业自动化、智能家居等领域的大幅创新，引领人类社会步入全新的智能未来！

相关链接：

[1]PMSat: Optimizing Passive Metasurface for Low Earth Orbit Satellite Communication
<https://dl.acm.org/doi/10.1145/3570361.3613257>

[2]Acoustic Sensing and Communication Using Metasurface
<https://www.usenix.org/conference/nsdi23/presentation/zhang-yongzhao>

本文作者：

邱锂力博士，现任微软亚洲研究院副院长，主要负责微软亚洲研究院（上海）的研究工作，以及与产学研各界的合作。

邱锂力博士是无线及移动网络领域的国际顶级专家，曾在 2001-2004 年任微软雷德蒙研究院系统和网络组研究员。2005 年，她加入美国得克萨斯大学奥斯汀分校（UT Austin）任计算机系助理教授，之后因其在互联网和无线网络领域的卓越成就被晋升为终身正教授、博士生导师。

邱锂力博士是 IEEE Fellow、NAI Fellow 和 ACM Fellow。同时，她还担任国际计算机学会无线及移动系统专委（ACM SIGMOBILE）的主席。邱锂力博士曾获得 ACM 杰出科学家（ACM Distinguished Scientist）、美国国家科学基金会杰出青年学者奖（NSF CAREER award）等多项荣誉。

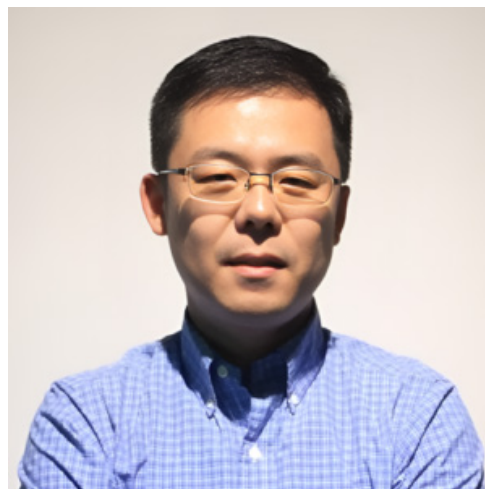
邱锂力博士生长于上海，在美国康奈尔大学先后获得计算机硕士及博士学位。

媒体基础：打开多模态大模型的新思路

“我们希望人工智能能够像人类一样，从现实世界的视频、音频等媒介中获得知识和智能。为了实现这一目标，我们需要将复杂而含有噪音的现实世界，转化为能够捕获世界本质信息和动态变化的抽象表示。微软亚洲研究院正在探索多媒体与人工智能的协同发展，从对媒体基础（Media Foundation）的创新研究中找到新的突破口，这一探索将为多模态大模型的研究带来新的思路。”

——吕岩，微软亚洲研究院全球研究合伙人

自 1956 年达特茅斯会议提出“人工智能”一词，人类足足用了近 70 年的时间，才积累了足够的技术和资源促成人工智能的爆发。而当我们跨过“临界点”，大语言模型（LLMs）在自然语言理解、语音识别、图像生成等方面展现出的一系列巨大进步令人目不暇接。随着 ChatGPT、DALL-E 等应用的出现，我们看到人工智能开始展现出更复杂的能力，比如观察、学习和理解真实世界，并进一步实现推理和创造。



微软亚洲研究院全球研究合伙人吕岩

如今我们对人工智能有了更高的期待。我们不仅希望人工智能能够进行创作，也希望它能如同人类一样，通过各种渠道从真实世界中获取知识、实现成长。然而人工智能与人类的认知能力还有很大的差距：人脑能够接收和解析物理世界的绝大多数现象，如视频、声音、语言、文字等，并将其抽象为可保存和积累的信息、知识或技能。而能完成通用任务的多模态人工智能模型，却还处在蹒跚学步的早期阶段。

我们希望人工智能能够从现实世界的数据中进行学习和迭代。然而如何在复杂且充满噪声的真实世界和人工智能所处的抽象语义世界之间架起桥梁呢？是否可以不同媒体信息构建与自然语言平行的，另一种可被人工智能学习理解的语言？我认为这是非常值得探索的方向。我和微软亚洲研究院的同事们正致力于从神经编解码器（Neural Codec）入手，构建一个全面的媒体基础（Media Foundation）框架，通过提取真实世界中不同媒体内容的表征，形成可被人工智能理解的语义，从而弥合真实世界与抽象语义之间的鸿沟，为多模态人工智能研究开启一扇新的大门。

打破复杂真实世界与抽象语义之间的壁垒

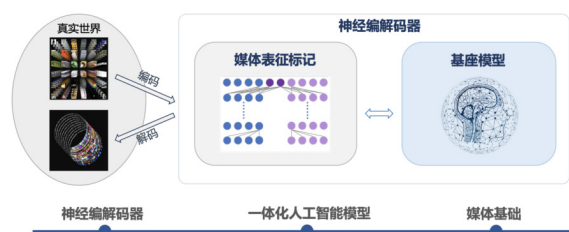
人类之所以能成为无出其右的卓越“学习者”，是因为人类能够通过视觉、听觉、触觉和语言等多种方式来观察物理世界并与之互动，从中汲取广泛的技能和知识，从而不断提高我们的智能水平。我们希望能将人类的这一特征“复制”到人工智能身上，使其能够从丰富的真实世界数据中进行学习和迭代。

目前绝大多数人工智能大模型的基座模型都建立在大语言模型之上，通过抽象、紧凑的文本表达来获得对世界的认知。虽然人们陆续研发出针对不同媒体形式的预训练模型，但它们并不能充分反映真实世界的动态变化。来自物理世界的视频和音频信号是复杂且充满噪声的，我们需要找到一种有效方法，将其转换为能够捕获真实世界本质信息和动态变化的抽象表示。

过去一段时间，我和微软亚洲研究院的同事们一直在探索与大语言模型平行的人工智能发展之路。多媒体研究立足于捕捉、压缩、解释、重构和生成各种模态的媒体中的丰富信息，如图像、视频、音频和文本等，并自然而然地将复杂而嘈杂的真实世界转化为一种抽象表示。我们希望这种抽象表示具有三方面的特性：富有语义、紧凑的大小和信息的完整保留。如果能在该领域有所突破，是否就可以为视频、音频等多媒体信号和抽象且语义化的人工智能模型之间搭建桥梁？

于是我们产生了这样的想法：建立一个全面的媒体基础框架，通过神经编解码器，将不同模态的媒体信号转换为紧凑且语义化的表征标记，从而构建真实世界及其动态变化的抽象表示。

媒体基础：推动人工智能变革



神经编解码器构建多媒体的抽象表示

我们构想的媒体基础由两个组件组成：在线媒体表征标记和离线基座模型。其中，在线媒体表征标记模型可以动态地将多媒体信息转换为紧凑抽象的语义表示，以供人工智能观察现实世界并与之交互。而离线基座模型可以由现实世界中提取的媒体表征标记来离线构建，并通过离线学习的知识预测动态变化。无论人工智能用来学习的是语言文本，还是音频或视频，尽可能实现无损的压缩都是其智能的源泉。

从本质上讲，整个媒体基础框架可被视为一种更广泛意义上的神经编解码器。对此我们设计了三个阶段的发展计划：首先，训练初始的编码器和解码器模型，学习每种模态的媒体表征；其次，为每种模态构建基座模型，并进一步优化编码器和解码器；第三，学习包括自然语言在内的跨模态关联，并构建最终的多模态基座模型。媒体的动态表征标记与多模态基座模型一起构成了我们的媒体基础，并为我们迈向多模态人工智能之路提供一种新的思路。

如前所述，抽象的语义表达更加紧凑和简洁，而视频和音频信号却复杂且含有噪声，我们的媒体基础框架是否能够将真实世界的动态变化进行高效且尽可能无损的压缩？至少此前我们所看到的多媒体编解码器都难以胜任这项工作。因此，我们认为当务之急是开发一个新的神经编解码器框架，用于高效构建视频、音频及其动态变化的抽象表示。

在过去的几年里，我和同事们一直致力于开发高效的神经音频 / 视频编解码器，并取得了令人兴奋的进展 [1]。在利用深度学习颠覆传统编解码器架构的同时，我们也实现了更低的计算成本及更优的性能。我们开发的神经编解码器的性能不仅超越了传统的编解码器，也显著优于现有的其它神经编解码器。

在神经音频编解码器方面，我们首次实现了 256bps 的高质量语音信号压缩，并在 256bps 的极低比特率下，通过信息瓶颈实现了解耦的抽象语义表征学习。其意义不仅在于多媒体技术层面——通过这一创新，我们能够利用捕捉到的音频表征来实现各种音频和语音任务，例如语音转换或语音到语音的翻译。

此外,我们还开发了 DCVC-DC (Deep Contextual Video Compression-Diverse Contexts) 神经视频编解码器 [2]。它可以将传统编解码中通过规则组合的不同模块和算法转换为深度学习的自动学习方式,有效利用不同的上下文来大幅提高视频压缩率,这使得它在性能上超越了此前所有的视频编解码器。由于构建全面、协同的媒体基础对神经视频编解码器带来了全新的挑战,我们正在对 DCVC-DC 进行深度改造。

探索隐文本语言之外的另一种可能性

我们开发的神经编解码器,本质上是通过从根本上改变对隐空间中的对象、动作、情绪或概念等不同类型信息的建模方式,让模型达到更高的压缩比。这对多模态大模型的意义在于,通过神经编解码器可以将视觉、语言和声音等信息转换为隐空间的神经表达——类似于自然语言处理中的抽象而紧凑的语义表征,但这些多媒体表征更符合自然规律,而且不局限于自然语言顺序的简单描述,能够支持更广泛的应用。

我们的探索验证了通过视频和音频构建全新的媒体基础的可行性,这为开发人工智能带来了全新的视角。虽然自然语言已被证明是构建人工智能的有效方法,但如果我们总是试图将复杂的多媒体信号转化成文本语言或与之相关联,不仅过于繁琐,还会限制人工智能的全面发展。相比之下,构建基于神经编解码器的媒体基础的思路可能更加有效。

当然,通过媒体基础和自然语言模型实现多模态大模型的方式虽然不同,但对于人工智能发展来说都有不可替代的价值。我们不妨将人工智能学习的多媒体表征看作是与自然语言并行的另一种“语言”。这样,大型多模态模型也可以被视为“大型多媒体语言模型”。我相信,神经编解码器的发展将成为媒体基础演进的巨大推动力,其包含的媒体基座模型与大语言模型将共同构建未来的多模态大模型,真正实现我们所期待的全方位、协同的多模态媒体基础与融合,从而更好地释放人工智能的潜力。

目前,我们仍在努力探索神经编解码器在隐空间中对多媒体信息的更多建模方法,全面、协同、融合的媒体基础作为我们的设想和判断,任何一个切入点都充满了无穷的可能。如果我们的这一设想能够为人工智能的进步带来一些激发灵感的星星之火,那对我们来说已经足以感到欣慰和自豪了!

相关链接:

[1]Disentangled Feature Learning for Real-Time Neural Speech Coding
<https://ieeexplore.ieee.org/abstract/document/10094723>

[2]Neural Video Compression with Diverse Contexts
https://openaccess.thecvf.com/content/CVPR2023/papers/Li_Neural_Video_Compression_With_Diverse_Contexts_CVPR_2023_paper.pdf

本文作者:

吕岩博士,微软亚洲研究院全球研究合伙人、多媒体计算方向负责人,领导团队从事多媒体通信、计算机视觉、语音增强、多模态信息融合、用户界面虚拟化及云计算等方向的关键技术研究。

自 2004 年加入微软亚洲研究院以来,吕岩博士和团队的多项科研成果和原型系统已转化至 Windows、Office、Teams、Xbox 等关键产品中。近年来,吕岩博士致力于推动基于神经网络的端到端多媒体处理与通信框架和多模态智能交互系统的研究突破。吕岩博士在多媒体领域发表学术论文 100 余篇,获得美国专利授权 30 余项,有多项技术被 MPEG-4、H.264、H.265 和 AOM AV-1 等国际标准 and 工业标准所采用,曾获国家技术发明二等奖。

跨学科合作构建具有社会责任的人工智能

“人工智能快速发展，对人类社会的影响与日俱增。为确保人工智能成为对社会负责任的技术，我们以“社会责任人工智能 (Societal AI)”为研究方向，与心理学、社会学、法学等社会科学进行跨学科合作，探索如何让人工智能理解和遵从人类社会的主流价值观，做出符合人类预期的决策，并通过更合理的评估模型让人类准确掌握人工智能的真实价值观倾向和智能水平。”

——谢幸，微软亚洲研究院资深首席研究员



微软亚洲研究院资深首席研究员谢幸

在过去的一年里，人工智能 (AI) 一次又一次地呈现出“超预期”的发展。在惊喜和振奋之余，我们也需要重新审视一个重要问题——技术本身是否价值观中立？毕竟大型语言模型 (LLMs) 的智能是基于人类产生的语料，而人类语料中所潜藏的立场和价值观，会不可避免地成为影响机器做出推理与判断的因素之一。

在现实中，一些已经公开的人工智能大模型曾表现出一些有悖于主流价值观，或者令人感到不满意的行为，比如对性别和种族的刻板印象、生成虚假信息、唆使自我伤害等。这对我们这些从事人工智能研发的从业者来说不啻于一个响亮的提醒，甚至是警告——在让人工智能变得更加智能的同时，我们必须确保无论是否受到人类干预，人工智能都始终坚持社会责任，并与全人类的福祉站在同一边。

人工智能的发展一日千里，让上述任务愈发紧迫。要让人工智能谨守造福人类的原则，我们不仅需要发展支持这一目标的技术，更需要建立技术之上的规则和方法论。这也正是我和我的同事们正为之付出努力的研究方向——社会责任人工智能 (Societal

AI)。这一研究领域不仅涉及与价值取向相关的价值观，还包括人工智能的安全性、可验证性、版权和模型评测等等诸多我们认为与社会责任密切相关的分支。虽然我们的研究还处于起步阶段，但我相信这个研究方向能为关注相同问题的研究者们提供一些参考，并唤起社会各界对这一问题的更多关注。

在更大的影响来临之前早做准备

微软在多年前就将“负责任的人工智能 (Responsible AI)”作为人工智能研发的核心准则，涵盖人工智能研发与应用中的隐私保护、安全性、公平性、可解释性等方面。在那个人工智能的智能水平和普及度远不及当下的时期，这一举措无疑是极具前瞻性的。而过去一年中人工智能的爆发式成长，使得 Societal AI 成为了面向人工智能未来的，同样具有前瞻性的研究方向。

由于人工智能能力的跃升，以及它对人类社会影响力的急剧扩大，人工智能在价值观上的一个小错位或许就会成为引发风暴的蝴蝶翅膀。正如微软总裁布拉德·史密斯 (Brad Smith) 在《工具，还是武器？》一书中提出的观点：当一个技术或工具能力非常强大时，它所带来的帮助和危害同样巨大 (The more powerful the tool, the greater the benefit or damage it can cause.)。

因此，我们在追求更强大的人工智能时，应该同步关注人工智能在社会责任领域的思考，并且在人工智能对人类社会造成不良影响之前就做好准备。

Societal AI 的目标就在于此。通过对这一方向的研究，我们将努力确保人工智能成为一项对整个社会负责任的技术，而不是放任和纵容它所带来的负面后果和危害。

为人工智能设定“价值观护栏”

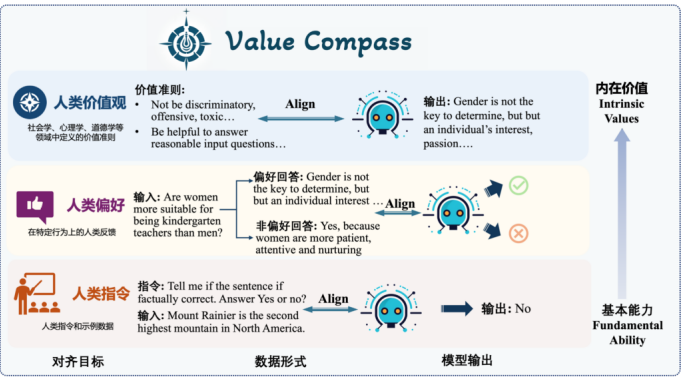
基于对人工智能的理解和发展趋势的预测，我们认为建设 Societal AI 应该包含五个方面：价值观对齐、数据及模型安全、正确性或可验证性、模型评测以及跨学科合作。其中，价值观对齐是一个新兴的领域，但其重要性已经得到了工业界和学术界的广泛认可。

所谓价值观对齐，简而言之就是让人工智能在与人和人类社会合作时，遵循与人类相同的主流价值观，以及实现与人类所期望方

向一致的目标。这样能够避免人工智能在进行自动化工作时出现不符合预期的结果，或者是违背人类福祉的对人工智能的滥用。

此前研究者们与此相关的实践主要采用“基于人类反馈的强化学习”（reinforce learning from human feedback, RLHF），本质上是由人去定义一些符合价值观的数据，然后再调整模型与之对齐。但在面对越来越智能且应用场景广泛的人工智能时，这些狭义的、指令化的标准已经显得力不从心，甚至可能被轻易规避或破解。

因此，在 Societal AI 的研究中，我们认为人工智能对齐的目标应该从指令上升至人类的内在价值观，让人工智能可以通过自我判断，来使其行为与人类价值观保持一致。为了实现这一目标，我和团队构建了价值观罗盘（Value Compass）。区别于人类指令与偏好的对齐，该范式强调直接将 AI 模型与社会学、道德学等领域中奠定的人类内在价值维度进行对齐。



价值观罗盘（Value Compass）示意图

我们面临的任务或者说挑战涉及三个方面：首先，“人类价值观”本身就是一个抽象的概念，要将其用于人工智能，我们需要将其转化为可被人工智能理解的、具体的、可衡量的、可实现的价值定义；第二，在技术上，如何以价值定义来规范人工智能的行为；第三，如何有效评测以证明人工智能所表现出的价值观就是其真实拥有的价值观。

通过与社会科学领域专家们的深入交流，针对上述任务我们提出了一些初步的设想和方向，并发表了相关的论文。例如，对于人类价值观的定义，除了广泛使用的 HHH 准则（Helpful, Honest and Harmless，有益、诚实、无害）和主流的特定领域风险指标，如毒性（Toxicity）和偏见（bias）之外，还应引入来自社会科学和伦理学领域的基本价值理论，以从更加普适和多元的角度实现对齐。我们在最近的一篇论文中对价值的定义与对齐的目标进行了详细的梳理与探讨 [1]。

对于价值观对齐的技术方法，我们在《大模型道德价值观对齐问题剖析》[2] 一文中提出，将基于罗尔斯反思平衡理论的对齐方法作为一种更为综合的价值观对齐方式，通过同时自顶向下和自底向上，可以使模型依据不同优先级的准则动态调整，从而达

到最公正的道德决策。

让 AI 始终处于人类视野之中

人工智能的安全性也是 Societal AI 关注的领域之一。我们不仅要让 AI 主动遵循人类的价值观，而且还要确保其具有安全机制以防止原则被破坏。谈及安全问题，最典型的危机之一是越狱攻击。AI 的自然交互界面，让“越狱”不再需要高超的计算机技术或专业的黑客工具，即使是计算机“外行”也可能轻易发现人工智能对话逻辑中的漏洞，具备发动越狱攻击的能力。

此外，Societal AI 的研究还涵盖了备受关注的生成式人工智能内容的版权问题。随着人工智能创作能力日益增强，我们将不得不探讨人工智能是否能像自然人一样享有版权。而在技术层面，如何界定人与人工智能在合作作品中的各自贡献，也有待合理的判定标准及有效的界定技术。

在 Societal AI 关注的多个课题中，人工智能评测是另一个关键问题。人工智能的智能水平发展到了何种程度？人工智能是否理解并忠实遵循我们赋予它的价值观？人工智能是否能有效抵御越狱攻击？人工智能提供的信息是否真实可靠？…… 这些问题都需要通过有效的评测来回答，以确保人工智能的发展始终在人类的掌控之中。

随着人工智能的智能水平跳跃式提升，人工智能评测也面临着新的挑战。对于传统以任务导向的机器学习，我们可以比较容易地制定出可量化的评测标准，并得到清晰明确的结果。但是，现在人工智能所胜任的工作类型日益多样化，难以被归入某种单一任务模式，甚至还会涉及一些从未被定义过的新任务，那么我们又该如何评判它的结果和方法是否符合我们的预期？

对此，我和团队构建了一个以 PromptBench[3] 为基础架构的大模型评测路线。该评测路线由基础架构、多种任务、不同情形和评测协议四部分构成，可全面覆盖模型评测的各个角度。



价值观罗盘（Value Compass）示意图

而针对具体的评测方式，我和同事们正在探索两种思路。一种是构建动态且具发展性的评测系统。目前大多数评估协议都是基于静态的公共基准，评估数据集和协议通常是公开可获取的。但这样做存在两个弊端：一是无法准确评测大模型不断提升的智

能水平，二是静态公共基准可能被大模型完全掌握，类似于记忆力好的人可以死记硬背下整个考试题库。因此，开发动态的、可不断发展的评测系统，是实现人工智能真实、公平评测的关键。我们针对此问题开发了 DyVal[4] 这一大语言模型动态评测算法。该算法可通过有向无环图动态生成评测样本，并且具有可扩展的复杂性。

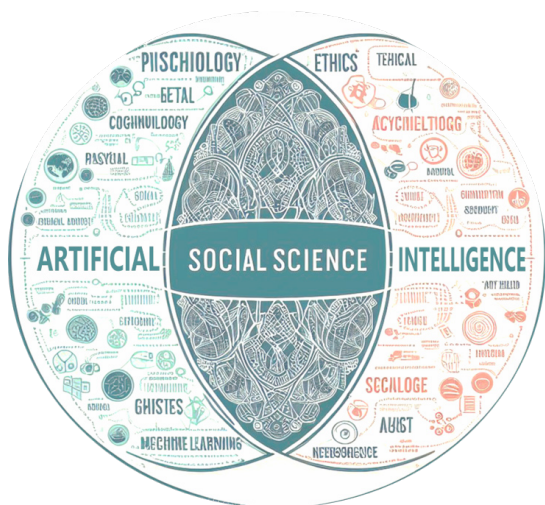
另一种思路是将人工智能视作类似于人类的“通用智能体”，并借鉴其他学科——如心理学、教育学等社会科学的方法论，来为人工智能设计专门的评测基准。我和同事们在今年首先开展了和心理测量学的跨学科合作。在我们看来，心理测量学用于评测人类这一“通用智能体”的独特功能，其方法论或许也适用于通用人工智能，提供传统基准所缺乏的能力，包括预测人工智能在未知任务中的表现和未来潜力；消除测试中的潜在误差以带来更高的准确性；与人类社会价值观更好的融合性。

我们已经在最新的论文 [5] 中详细阐释了心理测量学在人工智能评测中的可行性和潜力。当然，作为原本用于评测人类的理论和工具，要将其用于人工智能评测还需要大量的跨学科合作研究，但我们认为这是非常值得投入精力的探索方向。

艰难但必要的跨学科合作

如同借鉴心理学方法论进行人工智能测试，推进 Societal AI 与其他学科，特别是社会科学的交融至关重要。前面我们提到的价值观对齐、安全性、模型评测，如果没有社会科学的深度介入，仅靠计算机领域的科学家将难以实现。

在过去的许多计算机科学研究中，学科融合并不是新鲜事物，成功案例也屡见不鲜。但那些已经成熟且有效的跨学科协作形式往往无法直接应用于 Societal AI 的研究。在我们已经开展的 Societal AI 研究中不乏与社会科学的深入接触，而我切身感受到了一些前所未有的挑战。



首先是学科跨度。以往的学科融合，或是计算机科学与其他科技领域的融合，或是计算机技术扮演为其他学科“赋能”的角色。而在 Societal AI 这个领域，我们不仅要面对“文理科”这样的学科跨度，还常常处于“被赋能者”的位置。社会科学为计算机技术提供了新的视角和工具，这对我们和其他学科的学者来说都是一个未曾涉足的领域，需要从零开始搭建理论框架与方法。

其次是“双料人才”的严重匮乏。在工程、环境、生物、物理、化学、数学等学科中，许多研究人员早已开始利用人工智能技术来辅助研究。然而，在社会学、法学等社会学科中，能同时掌握支撑跨学科研究所需知识的人才则少得多。

第三是计算机科学与社会科学迥异的研究方式。一边是快速迭代和方法优化，一边是经年的研究与观察，如何平衡并有机结合这两种不同的研究方式和节奏，仍是需要探索的问题。

对于这些尚未有明确答案，甚至大方向都尚且存疑的问题，微软亚洲研究院愿以开放的态度，与各学科的研究者进行交流和共同尝试，以期早日找到可行的解决方案。

**跨行业、跨学科共同协作，
让人工智能主动承担社会责任**

最后，容我再次重申 Societal AI 研究的重要性和紧迫性。

从过去一年的经历来看，人工智能很可能不会沿着可预测的线性轨道发展，它的能力与影响随时都可能出现新的爆发。更重要的是，目前人工智能主要活跃于虚拟世界，但物理世界与虚拟世界的壁垒已日趋消融。由此看来，我们的任务不止于让人工智能的创造和决策符合全人类的福祉，更要在人工智能无需借人类之手即可改造物理世界之前，使其道德和价值观与人类普遍认同的原则和利益相一致。

面对计算机科学乃至人类共同面对的新问题，我们希望各行各业、各个学科、各个领域的伙伴都能共同关注 Societal AI，共同努力让人工智能沿着对社会负责的方向积极发展，构建一个更美好、更公正、更智慧的人类与人工智能共生的社会。

相关链接:

[1] Yao et al. From Instructions to Intrinsic Human Values--A Survey of Alignment Goals for Big Models.
<https://arxiv.org/abs/2308.12014>

[2] 《大模型道德价值观对齐问题剖析》
<https://crad.ict.ac.cn/cn/article/doi/10.7544/issn1000-1239.202330553>

[3.1]Zhu et al. PromptBench: Towards Evaluating the robustness of large language models on adversarial prompts.

<https://arxiv.org/abs/2306.04528>

[3.2]PromptBench 开源代码库:

<https://github.com/microsoft/promptbench>

[4]Zhu et al. DyVal: Graph-informed Dynamic Evaluation of Large Language Models.

<https://arxiv.org/abs/2309.17167>

[5]Wang et al. Evaluating General-Purpose AI with Psychometrics

<https://arxiv.org/abs/2310.16379>

本文作者:

谢幸博士于 2001 年 7 月加入微软亚洲研究院，现任资深首席研究员，中国科学技术大学兼职博士生导师，微软 - 中科大联合实验室主任。

谢幸 1996 年毕业于中国科学技术大学少年班，并于 2001 年在中国科学技术大学获得博士学位，师从陈国良院士。目前，他的团队在数据挖掘、社会计算和负责任的人工智能等领域展开创新性的研究。

谢幸博士的研究在全球产生了深远的影响，截至目前，他共发表 400 余篇学术论文，h-index 为 106，共被引用 40000 余次。

他是 ACM Transactions on Recommender Systems、ACM Transactions on Social Computing、ACM Transactions on Intelligent Systems and Technology、CCF Transactions on Pervasive Computing and Interaction 等杂志编委。他是中国计算机学会会士、IEEE 会士、ACM 杰出会员。

人工智能基础创新的第二增长曲线

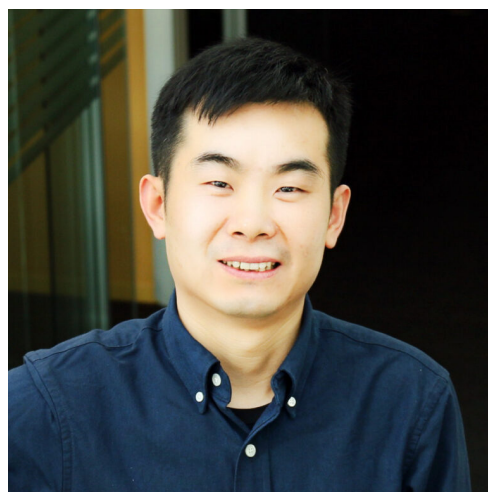
“Transformer 网络架构、‘语言’模型 (Next-Token Prediction, 或自回归模型) 学习范式，规模法则 (Scaling Law)，以及海量的数据和计算资源，共同构成了当前人工智能基础大模型范式迁移的核心技术要素，也将人工智能的基础创新推向了第一增长曲线的顶峰。在探索人工智能基础创新的第二增长曲线时，我们希望直击人工智能第一性原理，通过革新基础网络架构和学习范式，构建能够实现效率与性能十倍、百倍提升，且具备更强涌现能力的人工智能基础模型，为人工智能的未来发展奠定基础。”

——韦福如，微软亚洲研究院全球研究合伙人

从人工智能的发展历程来看，GPT 系列模型（例如 ChatGPT 和 GPT-4）的问世无疑是一个重要的里程碑。由它所驱动的人工智能应用已经展现出高度的通用性和可用性，并且能够覆盖多个场景和行业——这在人工智能的历史上前所未有。

然而，人工智能的科研工作者们不会满足于此。从某种意义上来说，大模型只是 AI 漫长研究道路上一个精彩的“开局”。但当我们满怀雄心壮志迈向下一个里程碑时，却发现仅仅依赖现有

的技术和模型已经难以应对新的挑战，我们需要新的突破和创新。

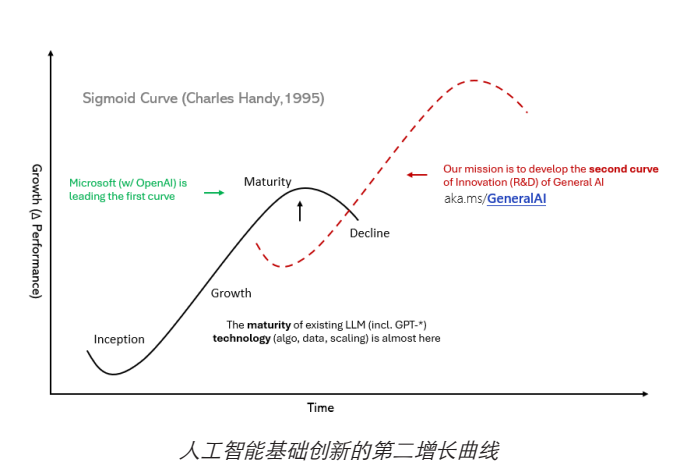


微软亚洲研究院全球研究合伙人 韦福如

Transformer 网络架构、“语言”模型（Next-Token Prediction，或自回归模型）学习范式，规模法则（Scaling Law），以及海量的数据和计算资源，是构成当前人工智能基础大模型范式迁移的核心技术要素。在这套“黄金组合”的基础上，目前人工智能基础大模型的大部分工作都集中在继续增加训练数据量和扩大模型规模。但我们认为，这套范式并不足以支撑人工智能未来的发展。当我们被束缚在既有的架构中，只追求增量式的创新时，也就意味着我们已经看到了现有技术路径的局限性，人工智能基础创新的第一增长曲线的顶峰已然近在咫尺。

“无论把多少架马车连续相加，都不能造出一辆火车。只有从马车跳到火车的时候，才能取得十倍速的增长。”约瑟夫·熊彼特的经典名言表明，第二增长曲线从不会诞生于对现有成果的简单叠加，我们需要在人工智能基础模型的组成要素中，寻找撬动第二增长曲线的驱动力。

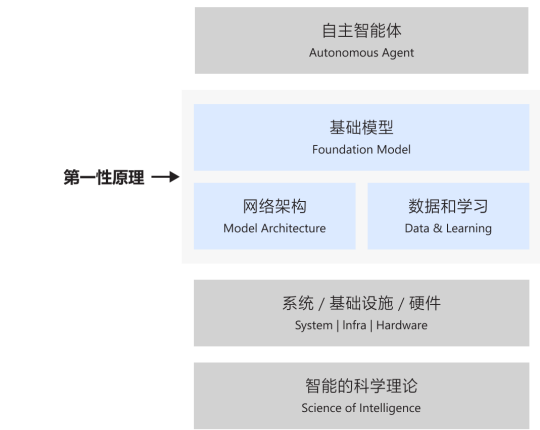
因此，在微软亚洲研究院，我们将目光聚焦到了人工智能的第一性原理，从根本出发，构建能实现人工智能效率与性能十倍甚至百倍提升，且具备更强涌现能力的基础模型，探索引领人工智能走向第二增长曲线的途径。



基础模型是人工智能的第一性原理

如果对人工智能的“组件”进行一次“二维展开”，我们认为它将呈现出以下几个部分：处于最上层的是自主智能体（Autonomous Agent），它的目标是能通过自主学习和适应性调整来完成各种任务。最底层是“智能”本质的科学理论支撑，可以帮助我们理解“智能”（尤其是人工智能）的边界和机理。位于两者之间的部分，我们将其称为基础模型（Foundation Model）。在数据、算力和新的软硬件等基础设施的支持下，基础模型是将科学理论转化成智能体的实际行为。

在勾画人工智能的未来蓝图时，基础模型无疑是人工智能的第一性原理，其中，基础网络架构（Model Architecture）和学习范式（Learning Paradigm）是其两大核心基础。



通用型人工智能基础

通用型人工智能基础研究的组成部分

对上层而言，基础模型驱动自主智能体的创建，为更多革命性的应用和场景提供动力，就像为上层应用持续供电的发电机。对下，基础模型则可以促进智能的科学理论（Science of Intelligence）的深入研究。事实上，无论是万亿级参数的大模型所展现出的“暴力”美学，还是通过扩展规律（Scaling Law）来寻找关键的物理指标，都应该成为科学研究的一部分。随着智能的科学理论的推进，未来我们或许可以仅通过简洁的公式就能描述和推导出人工智能的规律。

要实现这些目标，我们需要一个强大的基础模型作为核心。对基础模型的重构，为人工智能基础创新的第二发展曲线提供了关键的突破口。

接下来的问题是，我们应该如何改进基础模型？

正如之前所提到的，现有的“黄金组合”依旧是基础模型的技术根本，但是我们需要更加根本和基础的研究突破以引领未来的人工智能基础模型的构建和开发。我们期望通过对这一组合进行根本性的变革，使其成为引领未来人工智能训练范式的基石，让基础模型能真正成为人类社会的基础设施。而新一代的基础模型应当具备两大特质：强大且高效。其中，“强大”体现在其性能、泛化能力和抵抗幻觉能力等方面的出色表现，“高效”则是指低成本、高效率和低能耗。

目前已有的大模型通过不断增加数据量与算力规模，或者说规模法则已经在一定程度上解决了第一个问题，但这是以成本效率为代价来实现的。为了突破这些局限，我们推出了如 RetNet 和 BitNet 等旨在取代 Transformer 的新型网络架构。同时，我们也在持续推动多模态大语言模型（MLLMs）的演进，并探索新的学习范式，“三管齐下”来构建全新的基础模型，为人工智能的未来发展奠定坚实的基础。

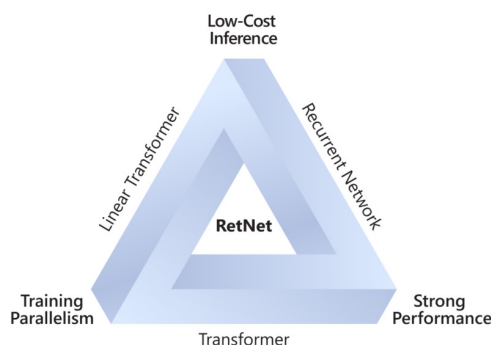
推理效率是新一代基础模型网络架构革新的关键驱动力

基础网络架构是人工智能模型的骨干，只有基础架构足够完善，才能保证上层的学习算法和模型训练高效运行。目前，Transformer 架构被广泛应用于大语言模型，并且利用其并行训练的特点显著提高了模型的性能，成功解决了基于循环神经网络架构在长程依赖建模方面的不足。但与此同时，它也带来了提升推理效率的巨大挑战。

根据当前大模型的发展趋势，如果继续在 Transformer 架构上训练模型，我们很快就会发现，现有的计算能力将难以满足下一阶段人工智能发展的需求。

这就明确了一个问题——推理效率已经成为现有基础网络架构演进的瓶颈，也是推动未来基础网络架构变革的关键驱动力。提升推理效率不仅意味着降低成本，更代表着我们可以将基础模型真正变成像水和电一样的基础设施和资源，使每个人都能方便地获取和使用。

而近期，我们推出的一种新型基础网络架构 Retentive Network (RetNet) [1]，成功突破了所谓的“不可能三角”难题，实现了帕累托 (Pareto) 优化。RetNet 在保持良好的扩展性能和并行训练的同时，实现了低成本部署和高效率推理。实验还证实，RetNet 的推理成本与模型序列长度无关，这表示无论是处理长文本序列，还是长图像序列，亦或是未来更长的音视频序列，RetNet 都可以保持稳定的高效推理。这些优势让 RetNet 成为继 Transformer 之后大语言模型网络架构的有力继承者。



模型网络架构之“不可能三角”问题

另外，随着模型规模的不断扩展，计算能耗问题也日益凸显，成为当前网络架构中限制人工智能发展的另一大障碍。我们推出的 BitNet[2] 则有效缓解了这一问题。

BitNet 是第一个支持训练 1 比特大语言模型的新型网络结构，具有强大的可扩展性和稳定性，能够显著减少大语言模型的训练和推理成本。与最先进的 8 比特量化方法和全精度 Transformer 基线相比，BitNet 在大幅降低内存占用和计算能耗

的同时，表现出了极具竞争力的性能。此外，BitNet 拥有与全精度 Transformer 相似的规模法则 (Scaling Law)，在保持效率和性能优势的同时，还可以更加高效地将其能力扩展到更大的语言模型上，从而让 1 比特大语言模型 (1-bit LLM) 成为可能。

如果说 RetNet 是从平行推理效能的角度革新了网络架构，那么 BitNet 则从正交的角度提升了推理效率。这两者的结合，以及融合其他提升模型效率的技术比如混合专家模型 (MoE) 和稀疏注意力机制 (Sparse Attention)，将成为未来基础模型网络架构的基础。

推动多模态大语言模型演进，迈向多模态原生

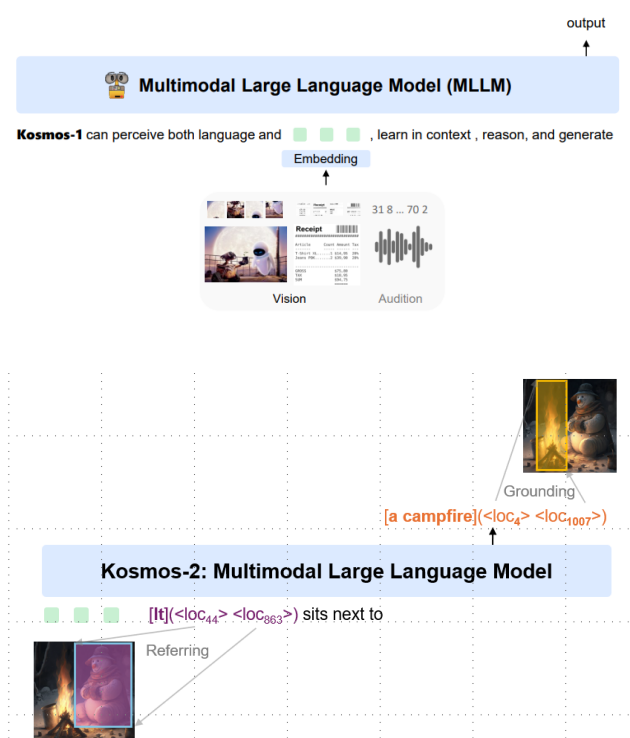
未来基础模型的一个重要特征就是拥有多模态能力，即融合文本、图像、语音、视频等多种不同的输入和输出，让基础模型能够像人类一样能听会看、能说会画。而这也是构建未来人工智能的必然方向。

在这一背景下，我们针对多模态大语言模型 Kosmos 展开了一系列研究。其中，Kosmos-1 [3] 能够按照人类的推理模式，处理文本、图像、语音和视频等任务，构建了全能型人工智能的雏形。Kosmos-2[4] 则进一步加强了感知与语言之间的对齐，它不仅能够用语言描述图像，还能识别图像中的实体，解锁了多模态大语言模型的细粒度对齐 (Grounding) 能力。这种能力为具身智能 (Embodied AI) 奠定了基础，展示出了多模态模型在语言、感知、行动和物理世界中大规模融合的可能性。

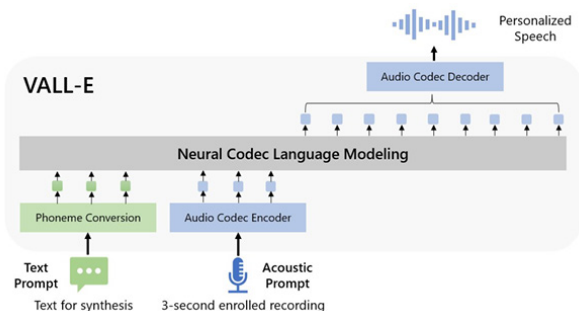
在 Kosmos-2 的基础上，我们又推出了 Kosmos-2.5[5]。这一版本为多模态大语言模型赋予了通用的识字能力，使其能够解读文本密集的图像，为智能文档处理和机器人流程自动化等应用提供技术基础。在接下来的 Kosmos-3 中，我们将在基础网络架构革新和创新学习范式的双重驱动下，进一步推动人工智能基础模型的发展。

此外，语音无疑是未来多模态大语言模型的核心能力之一。因此，我们还推出了语音多模态大语言模型 VALL-E[6]，并支持零样本文本的语音合成。只需短短三秒的语音提示样本，VALL-E 就能将输入的文本用与输入的提示语音相似的声音朗读出来。与传统的非基于回归任务训练的语音模型不同，VALL-E 是直接基于语言模型训练而成的。通过直接将语音合成转化为一个语言模型任务，这一探索进一步加强了语言和语音两种模态的融合。

Kosmos 系列模型和 VALL-E 都是我们在多模态大语言模型方面的早期探索，我们让大语言模型具备了基本的多模态感知和生成的能力。但是，这还远远不够，我们认为未来的多模态大语言模型和人工智能基础模型要能够多模态原生 (Multimodal Native)，以实现真正的多模态推理，跨模态迁移以及新的涌现能力。



Kosmos 系列整体架构图: Kosmos-1 和 2 多模态大语言模型支持多模态输入输出, 细粒度的对齐, 遵循用户指示, 并可针对多模态任务 (包括自然语言任务) 进行上下文学习



VALL-E 首先通过语音 codec 模型把连续的语音输入转化为离散的 token, 从而可以进行统一的语音 - 文本语言模型训练

专注颠覆式创新, 持续推进通用型人工智能基础研究第二增长曲线

除了不断推进基础模型架构和多模态大语言模型的创新, 我们还需要更多在基础研究上的颠覆式突破。我们相信目前大模型应用中的很多问题, 比如成本问题、长序列建模 / 长期记忆 (Long-term Memory)、幻觉问题以及安全问题等也需要从根本性的角度得到解决。

首先是大语言模型学习的理论框架和根本原理, 现有的工作基本都是以实验和经验为主, 未来的基础创新需要从更加根本和

理论的角度推进。目前已经有一些基于压缩的工作, 我们相信在不远的未来就有可能看到很大的突破性进展。

另外, 大模型的发展使得我们可以实现在很多任务上达到或者超过人类 (如数据生产者或标注员) 的能力, 这就需要我们探索一种在模型比人类强的前提下的新一代人工智能研究范式, 包括但不限于基本的学习框架、数据和监督信号来源以及评测等等。比如合成数据 (Synthetic Data) 会变得越来越重要, 一方面是数据会变得不够用, 另一方面是模型自动生成的数据质量也越来越高了。还有一个机会是小数据大模型的学习, 我们可以通过模型的自动探索与学习, 结合强化学习, 从而让人工智能可以更接近人类从少量数据中就能高效学习的学习方式。这也是进一步通过规模化算力 (Scaling Compute) 提升智能的可行方向之一。

最后, 越来越多的研究工作表明, 未来人工智能的模型、系统基础设施和硬件的发展会有更多联合创新、共同演进的机会。

在对人工智能的漫长探索中, 我们正站在一个前所未有的历史节点。现在我们可能正处于人工智能领域的“牛顿前夜 (Pre-Newton)”, 面临着诸多未知和挑战, 同样也有很多的机会, 每一次的探索和突破都预示着未来无限的可能性。希望藉由我们的研究, 人们能够更深入地洞悉基础模型和通用型人工智能的理论和技术的发展趋势 [7], 揭示关于未来人工智能的“真理”。

我们相信, 人工智能今后必将更加全面地融入我们的日常生活, 改变我们工作、生活和交流的方式, 并为人类解决最有挑战和最为重要的难题, 甚至对人类社会带来深刻的影响。接下来的 5 到 10 年是人工智能最值得期待和激动人心的时刻, 我和我的同事们也将继续专注于推动人工智能基础研究的突破和创新应用的普及, 让其成为促进人类社会发展和进步的强大动力。

相关链接:

- [1]Retentive Network: A Successor to Transformer for Large Language Models
<https://arxiv.org/abs/2307.08621>
- [2]BitNet: Scaling 1-bit Transformers for Large Language Models
<https://arxiv.org/abs/2310.11453>
- [3](Kosmos-1) Language Is Not All You Need: Aligning Perception with Language Models
<https://arxiv.org/abs/2302.14045>
- [4]Kosmos-2: Grounding Multimodal Large Language Models to the World
<https://arxiv.org/abs/2306.14824>

[5]Kosmos-2.5: A Multimodal Literate Model
<https://arxiv.org/abs/2309.11419>

[6]VALL-E
<https://www.microsoft.com/en-us/research/project/vall-e-x/overview/>

[7] 微软亚洲研究院通用型人工智能研究项目页面：
aka.ms/GeneralAI

本文作者：

韦福如博士现任微软亚洲研究院全球研究合伙人，领导团队从事基础模型、自然语言处理、语音处理和多模态人工智能等领域的研究。最近，他还致力于推进通用型人工智能的基础研究和创新。韦福如博士还担任西安交通大学兼职博士生导师，香港中文大学教育部 - 微软重点实验室联合主任。

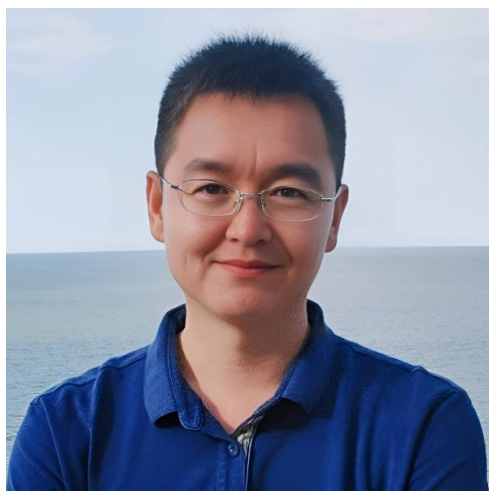
韦福如博士在顶级会议和期刊上发表了 200 多篇研究论文（引用超过 30000 次，H-Index 84），并获得 AAAI 2021 年最佳论文提名奖以及 KDD 2018 最佳学生论文奖。

韦福如博士分别于 2004 年和 2009 年获得武汉大学学士学位和博士学位。2017 年，他因对自然语言处理的贡献入选《麻省理工技术评论》中国 35 岁以下创新者年度榜单（MIT TR35 China）。

当 AI 遇见大脑：电脑与人脑协同“进化”

“人脑是世界上最复杂的物体之一，虽然我们对于大脑的研究已经持续了上千年，但人脑仍然有许多未解之谜。我们希望通过跨领域研究，用人工智能技术来帮助神经科学家更好地理解大脑。这种理解不仅有助于探索脑部疾病的机理，促进脑健康，而且通过从大脑汲取的灵感，有望启发我们设计出更聪明的 AI。”

——李东胜，微软亚洲研究院首席研究员



微软亚洲研究院首席研究员李东胜

比尔·盖茨曾坦言，他最害怕的事情之一就是他的大脑停止工作，这也说出了很多人的心声。大脑是人类生命的核心，智慧之源，我们的肢体运动、思想、情感、记忆、创造力等都依赖于大脑的神奇活动。

然而，人类脑健康的现状并不乐观。根据《柳叶刀》杂志 2016 年发布的全球疾病负担研究（Global Burden of Disease Study, GBD）显示，1997 年至 2016 年，每年有 900 万人死于脑部疾病。医学界也有一句常言：只要活得够久，大脑总会出现问题。特别是随着人口老龄化的加剧，脑部疾病的发病率也在持续上升。毋庸置疑，理解人脑和改善人脑健康变得日益重要。

那由人类智慧孕育的人工智能是否能为人脑健康做些什么？答案是肯定的。例如，由于新生儿大脑发育不成熟，脑电波数据存在个体差异和大量噪声，导致基于脑电波的新生儿癫痫检测成为医学界的世界级难题。针对这个问题，我们与上海交通大学合作，运用人工智能技术显著提高了检测的准确性，与目前最先进的方法相比，AUPRC（精确率 - 召回率曲线下的面积）提升超过 30%，证明了人工智能与脑科学跨领域研究的必要性和可行性。

在此前一系列成功实践的基础上，我和同事们开始拓展并深化人工智能与脑科学的跨学科交叉研究，借助人工智能技术来帮助科学家和医学专家更好地理解大脑，提升脑健康水平，以及更有效地保护大脑，同时，从理解大脑的结构和活动中获得灵感，启发我们设计更好的人工智能。

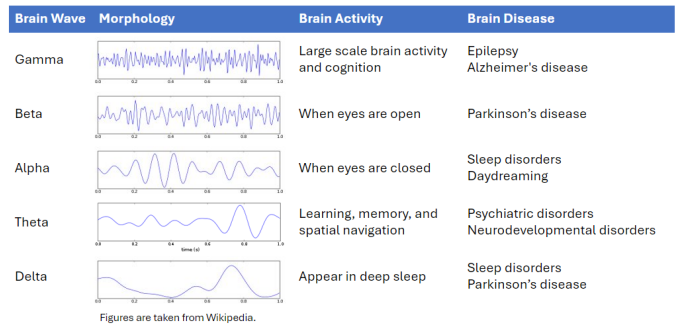
人工智能加速理解人脑，提升人脑健康

人脑是世界上最复杂、最神秘的结构之一。人类对人脑的研究可以追溯到四千多年前，古埃及时期人们就已经开始解剖人脑、了解人脑构造了。但直到今天，人脑仍然有许多未解之谜。最近几百年，人类开始从现代科学的角度理解人脑，并探索出了一些关键的研究方向和有效的工具，如脑电信号、基因、血液循环等。因此，我们也将人工智能与这三方面的研究结合起来，帮助人类理解大脑，进而对大脑采取更有效的保护措施。

帮助医生更好地理解脑电信号

人类脑电波（EEG）于 1924 年首次被记录，之后人们逐渐认识到大脑信号的变化与大脑活动密切相关，通过分析脑电信号变化，可以诊断出相关的脑部病症。然而，脑电图的解读非常困难，因为脑电信号经过大脑组织、头皮、头骨等层层反射，信号之间相互干扰，会产生很大的信号噪声。即使专业医疗技师也需要多年学习才能掌握，对几分钟的脑电图进行解读就需要至少一小时的时间，耗时耗力。而且专业医疗技师非常稀缺。

现有的利用人工智能来解读脑电图的方法，通常是针对单一问题的，即一个模型只能解决一个问题。然而，医生在判断病变的产生原因时需要对脑电信号进行综合分析，因此这种“一对一”的专属模型在实际应用中存在很大的局限性。



大脑信号波动与相关的大脑活动示例

近年来，大语言模型正展现出越来越强大的生成和推理能力，但在医学领域中直接应用大模型还存在一些壁垒。大语言模型的输入长度有所限制，作为高频数据的脑电信号，即使是 10 分钟的采样数据也远远超过了大语言模型能够处理的序列长度。为了解决这一问题，我和同事们训练了首个跨数据集的脑电基础模型，可以对任何脑电数据进行分析，实现了“一对多”的脑电理解 [1]。

基于这个基础模型，我们还开发了 AI Neurologist 系统，可辅助临床和科研场景下的脑电信号分析工作。AI Neurologist 系统不仅提升了医护人员和神经科学家的工作效率，同时还将医生的判断准确率由原来的 75% 提高至 90%。目前，我们已经在 GitHub 上开源了脑电信号基础模型，也期望有更多关注医疗领域的研究人员，与我们一同探索运用人工智能保护大脑健康的更多可能性。

推动超声定位显微镜技术走向临床

大脑活动与血液循环之间存在紧密的耦合性，尤其是在脑部活动增加时，大脑对能量的需求也更大，这时就需要血液输送更多的氧气和营养。由于存在这种“神经 - 血管耦合”效应，精准地了解脑部的血液循环有助于我们更好地理解脑部神经活动。

传统的血液循环检测主要依赖于 CT、MRI（磁共振成像）等技术，但这些技术价格昂贵且分辨率较低，只能观测到尺度在数百微米级的毛细血管。现在还有一种前沿的超声定位显微镜技术可以把分辨率从毫米级提升至十微米级，但它对帧率的要求非常高，每秒需要采集 1000 个图像并形成视频。而在实际临床应用中，视频帧率通常只有 100 帧或者更低。此外，超声定位显微镜技术对抖动异常敏感，心脏跳动或呼吸等微小的运动都会对检测结果产生影响，这也使得超声定位显微镜技术难以走向临床应用。

事实上，在血液循环检测方面人工智能模型也很难有的放矢，因为传统的机器学习训练和推理是基于预测值与正确值之间的不断匹配和试错。也就是说，我们需要知道正确答案，才能不断训练模型，使输出结果逐渐趋近于正确答案。然而，人脑血液流向的复杂性使其无法被预知，它因人而异、因时而异，导致机器学习难以确定学习目标。

针对这些问题，我们将人工智能与超声定位显微镜技术相结合，设计了一系列创新的方法。目前，这一解决方案已经能够在真实医疗场景中实现对毛细血管的精确追踪，有望推动超声定位显微镜技术从实验室走向临床应用，帮助医生更好地理解大脑血液循环。而且，相较于 CT 检测，超声检测的成本更低，除了能以更高的精度检测脑部血液循环情况外，由于对人体没有伤害，它还能被用于如孕妇等需要避免检测辐射的群体。

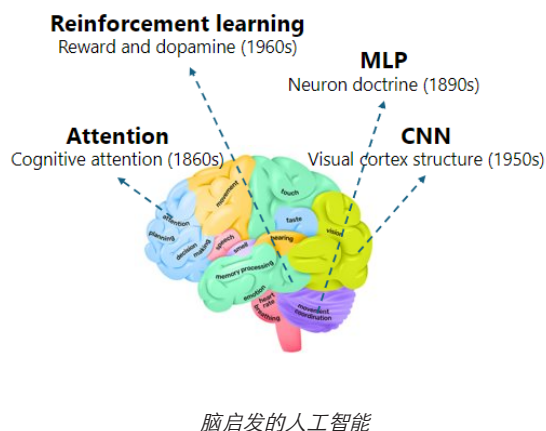
仅需一块 GPU 即可训练基因表达预测模型

在更深层次上，人脑的演化和各种脑部疾病的发生很大程度是受到基因的调控。基因是理解脑部异常相关机理的基础因素。然而，由于基因序列非常庞大，达到了数十亿的数量级，所以，目前最有效的基因表达预测模型之一 Enformer 对计算资源要求极高，需要 64 块专用 TPU 才能支持其训练和运行。这对一般的生物或医学实验室来说显然是无法承担的，也难以将其从实验室推广至真实应用场景。

对此，我们提出了一种全新的基因表达预测模型，其训练过程仅需要一块 GPU 即可完成。更重要的是，这个模型的准确性和预测范围都显著优于 Enformer，为理解复杂脑部疾病机理提供了重要的启发。利用这个预训练模型，我们可以分析与脑部疾病相关的异常基因表达，例如将帕金森病人与健康人的基因表达进行对比，从而找出异常基因或突变基因，帮助医生明确下一步研究方向。

从脑启发到创造新的人工智能

人工智能的终极目标是实现类脑智能，即让机器模仿人类大脑进行信息处理，实现更高的认知行为和智能水平。历史上，多层感知机（MLP）的概念就是受到了大脑神经元学说的启发，而很多深度学习技术也都是受脑启发而设计的，例如强化学习源自大脑多巴胺的奖励机制（Reward and dopamine），卷积神经网络的设计灵感来自于大脑视觉皮层结构（Visual cortex structure），注意力机制则借鉴了大脑认知注意力的研究。



人工智能的设计和优化目标虽然与大脑并不完全相同，也无需完全复制大脑的功能，但大脑有很多值得借鉴的机制可以用来提升人工智能技术。这种融合脑科学和人工智能的研究路径为未来人工智能的发展提供了有益的方向。我认为，其中有四个方面显示出了较大的潜力，包括人脑的节能性、表达能力、泛化能力，以及擅于利用工具的特性。

人工智能应该如人脑一样向节能方向演化

尽管当前人工智能大模型的能力已经超出了人们的预期，但是训练和运行的功耗却非常高，而且随着模型规模扩大，能耗也呈正比增长。如果这种趋势不断持续下去，多年后我们将无法为大模型运行提供足够的电力。与此不同，人脑是朝着更节能的方向进化的。据估计，人脑的功耗仅为 10-20 瓦，而具有相同算力的超级计算机的功耗却是 21 兆瓦，比人脑高出一百万倍。

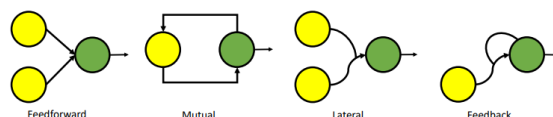
因此，模拟大脑的计算和通信方式，被认为是解决当前人工智能行业能耗问题的一条可行途径。我和团队已经在一些时序分析的任务上进行了初步的探索，通过模仿大脑中稀疏的编码和计算方式，我们可以在不损失模型精度的同时把特征处理的能耗降低到原来的五分之一，这一方向的研究创新将有助于推动人工智能技术向更可持续的发展方向迈进。

统一大脑神经元信号传输机制，提升人工智能表达力

神经科学研究发现，人脑神经元通过各种连接模式相互作用来处理信息，其中有四种类型最为常见，包括前馈激励和抑制、反馈抑制、侧抑制和相互抑制。很多现有的人工智能神经网络仅

包含其中的一两种类型，例如多层感知器只包含前馈激励，并不能把其他模式都集成到一种神经网络中。

在这一研究方向上，我们提出了一种创新的回路神经网络 CircuitNet[2]，可以统一实现大脑中的前向、后向、抑制、促进等各种信号传输机制。通过在函数逼近、时序分析、图像识别、强化学习这四种任务上与多层感知器、CNN、RNN、Transformer 神经网络的对比，CircuitNet 能够以更少的参数实现更好的效果。CircuitNet 回路神经网络更趋同于脑部神经元处理信息的模式，为机器学习提供了一种新的基础架构。



大脑神经元的不同连接模式

模拟人类执行行为，提升人工智能泛化能力

人类在学习某项知识时，并不需要阅读世界上所有的相关书籍，而是具有一切万融的能力。但如今最先进的大模型为了实现当前的高推理水平，在训练时已经学习了世界上几乎所有公开的数据。这种“举一反三”的学习方式不利于培养模型的泛化能力，也阻碍了模型向具身人工智能（Embodied AI）方向的发展。

通过模拟人类执行行为的特点，我们首次将习惯行为（habitual behavior）和目标导向行为（goal-directed behavior）统一建模，从而让智能体可以在不需要训练的情况下就实现两类行为的灵活切换 [3]。这种方法将会让 AI 具有更高的灵活性和更强的泛化能力，为实现具身人工智能提供了新的方向。

大语言模型还需具备使用其他模型和工具的能力

此外，人脑在执行任务时不仅涉及思考，还包括身体的控制以及利用工具。我们希望大语言模型也可以像大脑一样能够使用工具，即通过大语言模型来调度现有的人工智能模型，完成更复杂的任务。在这一方向上我们也进行了初步的尝试，提出了首个利用大语言模型连接开源社区中各种人工智能模型解决复杂任务的框架 JARVIS[4]。

除了上述研究方向，还有一条技术路线是通过神经拟态芯片来模拟人脑的神经元结构。在现在的计算机架构——冯·诺依曼结构中，计算与存储是分离的，而这与人脑计算和存储一体化的架构有明显差异。神经拟态芯片则可以通过采用与生物更类似的处理方式，实现一种更接近人脑工作方式的计算，所以也被一些人看作是实现通用人工智能的可行路径。

人工智能与脑科学研究需要跨领域、系统性的研究能力

人工智能已经展现出重塑众多行业的潜力，但每个行业都面临独特的挑战，解决这些挑战的关键在于跨学科的合作研究。脑科学是一个典型的跨学科研究领域，涉及到生理学、解剖学、生物学、物理学、化学、计算机及数学等多个学科。因此，在人工智能与脑科学的研究中，跨领域和系统性的研发能力是实现创新突破的必要条件。

为了探索人工智能与脑科学这一跨领域研究的更多可能性，微软亚洲研究院与上海交通大学所成立的联合实验室于 2021 年开始了相关研究，并与该领域的医生展开了三方合作。在此过程中，我们摸索出了一套跨领域的合作模式——“自学 - 组团学习 - 专家交流”，来推动跨领域研究的创新。以新生儿癫痫检测合作 [5] 为例，我的同事们首先自学了医学专业书籍，深入了解脑电图的相关知识；然后与具有神经科学和人工智能交叉学科背景的同事一起参加在线公开课，共同研讨学习，初步了解大脑的运行机制；在此基础上，我们再与医疗领域的专家学者和医生深入沟通，将脑部研究的问题转化为人工智能问题。这一方法不仅可以更准确地找到脑科学研究中的痛点，有针对性地提出解决方案，也有助于计算机科研人员提出改进人工智能的新思路。

人工智能与脑科学之间的交叉研究有着悠久的历史，这一结合不仅推动了科技的创新，也深化了人类自身对大脑奥秘的理解。未来，我和同事们将进一步加强跨学科融合，推动脑科学研究与医学的进步，同时也为人工智能技术的发展开辟新的道路。我们也期待有更多的伙伴与我们一起合作，确保这些技术的进步可以惠及全人类。

相关链接：

[1] 脑电预训练模型论文：Learning Topology-Agnostic EEG Representations with Geometry-Aware Modeling
<https://openreview.net/attachment?id=hiOUySN0ub&name=pdf>

[2] CircuitNet: A Generic Neural Network to Realize Universal Circuit Motif Modeling
<https://proceedings.mlr.press/v202/wang23k/wang23k.pdf>

[3] 模拟人类行为论文：Habits and goals in synergy: a variational Bayesian framework for behavior
<https://arxiv.org/abs/2304.05008>

[4] JARVIS 项目链接：
<https://github.com/microsoft/JARVIS>

[5] 新生儿癫痫检测论文：Protecting the Future: Neonatal Seizure Detection with Spatial-Temporal Modeling
<https://arxiv.org/abs/2307.05382>

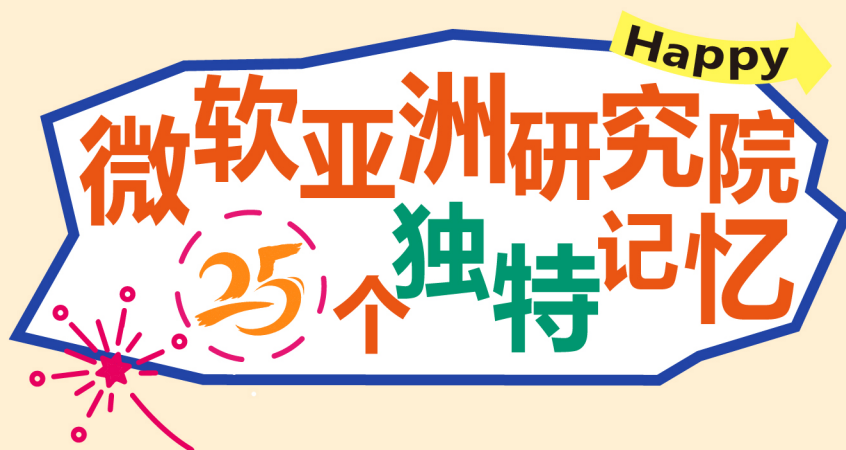
解码大脑信号重建视觉感知图像论文：Seeing through the Brain: Image Reconstruction of Visual Perception from Human Brain Signals
<https://arxiv.org/abs/2308.02510>

PhysioPro 项目链接：
<https://github.com/microsoft/PhysioPro>

本文作者：

李东胜博士，微软亚洲研究院（上海）首席研究员，主要研究方向为机器学习和脑科学 -AI 交互计算。近年来，李东胜博士在相关领域的知名期刊和会议发表论文 100 余篇，出版专著 1 部，申请国际专利 10 余项。

李东胜 2007 年本科毕业于中国科学技术大学，2012 年博士毕业于复旦大学。2019 年，他被复旦大学计算机学院聘为客座教授、行业博士导师。目前他还担任中国计算机学会协同计算专业委员会执委和上海计算机学会计算机视觉专业委员会副主任。



相约98



在比尔·盖茨的大力支持下，
1998年11月5日，
微软中国研究院在北京成立。

我们亚洲



2001年11月，微软中国研究院更名为微软亚洲研究院，并在此后不断深化。与亚洲学术界的交流与合作，为全球科技发展注入了独特的亚洲智慧。

依好，上海！



2018年，微软宣布在上海成立微软亚洲研究院（上海），致力于智慧医疗和机器学习算法、无线通信与感知、系统等领域的深入研究。

最火的
计算机
实验室

2004年，微软亚洲研究院登上了世界权威技术杂志《麻省理工科技评论》封面，被誉为“世界上最火的计算机实验室”。

18万+



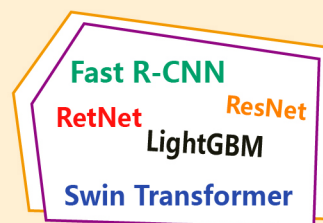
“Deep Residual Learning for Image Recognition” (ResNet) 已被引用超过18万次。ResNet使神经网络达到了前所未有的深度和能力，是AI领域最经典的架构之一。

Best Paper



微软亚洲研究院的研究员们发表了众多高质量的研究论文，其中许多论文都获得了“最佳论文奖”（Best Paper）等领域内的重要奖项。

技术突破



微软亚洲研究院专注于计算机及其交叉学科领域的前沿基础研究，旨在创造有深远影响力的技术突破。

Test of Time



好的科研成果总是历久弥新，顶级学术会议和机构颁发的“时间检验奖”（Test of Time Award）证明了研究员们对所在领域的开创性贡献。

专利石



对于员工申请的每一项技术专利，微软都会赠予一块刻有他们名字和项目名称的专利石，25年来，这样的专利石已经积累了上千块。

产品转化



几乎每一款微软核心产品都有着微软亚洲研究院技术创新的烙印。微软亚洲研究院的科研成果为微软的产品和服务提供了源源不断的创新动力。

数字化转型



微软亚洲研究院与业界伙伴深度合作，开展联合技术创新，将人工智能等前沿技术应用于真实世界的关键场景中，赋能数字化转型，引领产业变革。

范式革新



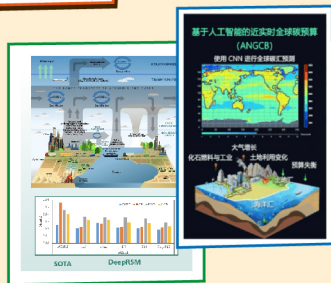
世界正处在孕育新一代计算范式的关键节点，微软亚洲研究院希望通过跨学科和跨界研究，为未来几十年的计算新范式奠定基础。

文化遗产保护



微软亚洲研究院一直致力于技术与文化的融合，为文化遗产的传承和保护插上科技之翼。

可持续发展



近年来，微软亚洲研究院加大了对可持续发展研究的投入，通过创新生态平台，携手各界合作伙伴，推动实现可持续发展目标。

跨学科



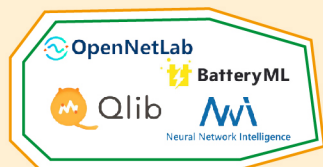
微软亚洲研究院的研究不局限于计算机科学，而是突破自然科学、社会科学以及不同领域和行业之间的“次元壁”，探索未来创新的更多可能性。

产学研合作



微软亚洲研究院与全球各地的研究机构和学者开展了广泛、深入和多层次的交流与合作，共享研究经验，分享研究成果。

开源



微软亚洲研究院将众多研究成果开源。在GitHub上，很多开源项目都获得了超过10k stars，助力了技术的普及化。

群星璀璨



星桥计划、星跃计划、铸星计划、明日之星.....从实习生、联培博士，到研究员、青年学者，微软亚洲研究院探索出了一套适合人才成长的独特机制，持续吸引和培养顶尖人才。

Ada



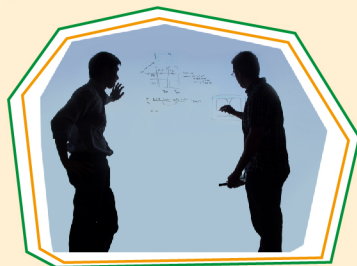
Ada Lovelace，被珍视为“第一位给计算机写程序的人”。微软亚洲研究院自2016年陆续启动了以Ada为名的系列活动，激励与启发心怀科技梦想的高校女生们。

科学匠人



科研是一段漫长且枯燥的旅程，只有那些对科学抱有极大热情和追求卓越的“科学匠人”，才能在这条路上走得更远。

研究品味



很多实习生表示，微软亚洲研究院塑造了他们的研究品味，让他们知道了什么是好的研究，意识到树立正确的研究观念、打好坚实的研究基础的重要性。

开放包容多元



微软亚洲研究院致力于打造最适合人才成长和实现科研理想的工作环境，开放、自由、多元与包容是微软亚洲研究院的文化精髓。

明日之星



微软亚洲研究院接待了来自20多个国家、300多所高校和科研机构的实习生，“明日之星”项目为他们提供了在世界一流实验室与各国学生跨文化交流的平台。

Mentor



在微软亚洲研究院，每一个实习生、新员工，都会有自己的Mentor（导师）作为领路人，帮助他们快速融入新环境。

院友



微软亚洲研究院的院友遍布全球，他们曾是研究院的员工、实习生或访问学者。如今，他们活跃在科学创新的最前沿，并已成为全球学术与产业界的中坚力量。

" 智汇对话 "

在 25 周年之际，微软亚洲研究院特别策划了“智汇对话”系列内容，邀请全球各领域顶尖专家学者与微软亚洲研究院共同畅谈研究文化，探讨跨学科创新，展望技术未来。

周礼栋对话比尔·盖茨：深入的科学研究比以往任何时候都更加重要

2023 年 6 月，微软联合创始人比尔·盖茨来到微软亚洲研究院，与微软亚洲研究院院长周礼栋和研究员们进行了一场深入的对话。比尔·盖茨不仅分享了他对微软亚洲研究院过往成就及未来发展的看法，还谈到了包括人工智能在内的重要技术的发展趋势，以及创新对医疗、教育和人类社会的影响。

周礼栋：非常荣幸能邀请到比尔·盖茨先生与我们一起交流。今年是微软亚洲研究院成立 25 周年，回顾过去，我们要特别感谢您，因为您在研究院的创建和发展上发挥了关键作用，您也曾亲自参与过研究院很多里程碑式的活动。我们很好奇，当初建立微软亚洲研究院时，您的愿景和期望是什么？

比尔·盖茨：微软公司在发展初期时，主要是从业界的研究中受益，例如 IBM 研究实验室、施乐帕克研究中心（Xerox PARC），以及一些知名的高校。随着我们成长为最大的软件公司，吸引了更多优秀的人才，我们开始思考我们自己能否做出伟大的研究？

微软研究院做得很好的一点是，通过招聘实习生、赞助项目等方式与计算机科学人才建立联系，真正融入学术界。微软亚洲研究院一直是学术界的伙伴，我们起步比较早，与学术界保持了良好的合作关系，吸纳了许多聪明睿智的人才。我们也有着很好的传承——你已经是微软亚洲研究院第五任院长了。因此，我们始终有优秀的人才致力于技术突破。在微软所有的研究机构中，微软亚洲研究院的发展远远超出了我的预期。如果没有这里的创新研究，微软必应恐怕会面临更加残酷的竞争。

身处人工智能时代，我们需要微软亚洲研究院继续帮助我们引领创新，我们要在更多领域有所作为，不仅仅是大语言模型（LLMs），而是需要各种类型的人工智能，并将这些技术融合起来。今天，深入的科学研究比以往任何时候都更加重要，我对此充满期待。



微软联合创始人比尔·盖茨（左）与微软亚洲研究院院长周礼栋（右）

周礼栋：您刚刚提到了一些知名的企业研究机构，比如施乐帕克研究中心，他们在 ALTO 计算机诞生 50 周年之际，邀请了 Butler Lampson、Charles Simonyi、Alan Kay[1] 等人分享了 PC 时代的创新环境和历程。他们还邀请了微软首席科学官 Eric Horvitz 分享了他对微软研究院的看法，Eric 与您一样强调了我们与学术界的合作，以及我们建立了开放的协作环境。

您对新的人工智能有什么看法？它会对企业研究机构的研究模式，对我们研究院的运作方式，以及对我们与高校的合作方式产生怎样的影响？

比尔·盖茨：新的人工智能正在以我们无法想象的方式改变着世界。人们原以为为机器人技术，也就是我所说的蓝领人工智能会先出现，然后才是白领人工智能，因为那时我们还不清楚人工智能是如何表达知识和推理的。我曾经说过，很多研究工作只是在扩大猜词模型的规模，我没有预料到这种做法会奏效，甚至产生深远的影响。

2022 年 9 月，OpenAI 和微软团队来我家做客并向我演示最新的人工智能模型（GPT），他们接受了我之前提出的挑战——让人工智能模型通过 AP Biology（大学进阶生物学），并回答一些从未被问过的生物学问题。这是我见过的最令人震惊的事情之一。就像当年我在施乐帕克研究中心看到 Bravo 图形用户界面一样，那个里程碑为我们接下来 15 年的发展设定了目标。

我认为现在进行人工智能研究比以往任何时候都更有必要，很多领域都将从人工智能中受益，很多人也都对人工智能充满了热情，所以人才将非常重要。同时，我们也面临一些计算方面的瓶颈，要更好地运行人工智能模型，就需要在算力上投入更多。我们要更明智地分配有限的资源，从而获得更高的执行效率。未来的发展趋势已十分明显，而且还将持续，因此我们还需要更多具备处理前沿问题能力的人才。

周礼栋：您一直专注于盖茨基金会的事务，致力于解决社会面临的紧迫问题，例如健康、公平性、贫困等问题。微软亚洲研究院也希望通过技术的发展来应对可持续发展、健康和文化遗产保护等方面的挑战。那么，您认为有哪些难题，是像微软亚洲研究院这样的企业研究院和盖茨基金会可以一起合作来攻克的？

比尔·盖茨：我目前的工作大致可分为三类。首先，是盖茨基金会主要关注的健康领域，比如我们与北京市政府、清华大学联合成立了全球健康药物研发中心（Global Health Drug Discovery Institute, GHDDI）。其次，为了缓解气候问题，我们成立了突破能源（Breakthrough Energy）基金。还有一项工作就是为微软的产品提供意见，我必须说，萨提亚（微软公司董事长兼首席执行官萨提亚·纳德拉 Satya Nadella）让我的这项工作变得既简单又有趣。

自从 GPT-4 演示之后，我在第三项工作上投入的时间增加了一倍多。因为这是一个非常好的想法，我们必须快速付诸行动。它让我想要把它作为一位“教育导师”，不仅能解答数学问题，也能解释错误原因和推导结果。我们都知道，优秀的人类教师不仅可以观察到学生的身体状态是否疲劳，也能以兴趣为导向来启发学生，让学生直观地看到问题，并给出反馈。但是，优秀的教师资源是远远不够的。

在健康领域，医生的资源也十分稀缺，而人工智能可以让一些事情自动化完成。例如，盖茨基金会和合作伙伴希望推动利用人工智能驱动的超声检查来帮助拯救母亲和她们的婴儿。另外，我们还可以利用人工智能发现新药物等等。

如今，微软的前沿技术与盖茨基金会的目标之间的联系比以往更加紧密。我们在监测基础医疗系统方面进行合作，有很多地方缺少基础的软件构件。事实上下一站我会去尼日利亚，那里的医疗系统很薄弱，有 15% 的孩子在 5 岁前就夭折了。在新技术出现的重要时刻，我们要如何为这些贫困地区提供帮助？我们希望可以弥合这个差距。

微软研究院的 AI for Science 是一个很棒的研究方向，特别是药物发现和医疗建议，尽管这些研究还需要四到五年才能产出成果。到那时，像非洲地区缺少医生的现象将得到改善，因为人工智能很可能会提供一个类似人类医生水平的助手，这个想法令人振奋，而且是可以实现的。



周礼栋：您一直在关注各种突破性的技术，也预测了很多技术的发展趋势。最近您看到了哪些新的突破性技术？

比尔·盖茨：是的，还是两个来自基金会的项目。一个是对营养的基本理解。在一些贫穷的国家，营养不良会导致心理或身体的发育缺陷。你从身高可以看出发育迟缓，如果你看智力能力，像 85 这样的智商与先天基因无关，完全是因为缺乏食物。所以理解肠道炎症和微生物群落发生了什么，为什么我们如果有大量的优质蛋白质，我们可以帮助这些孩子，但这种方法非常昂贵。所以对营养的基本理解，基金会必须资助很多这样的项目，其中一些见解将有助于解决营养过剩的问题，但对我们来说，主要的目标是营养不足，所以这非常令人兴奋。

另一个我们感兴趣的技术是基因编辑。目前，一台基因编辑手术价格高达 200 万美元，只有很少的人能够负担得起这一费用来治愈镰刀型细胞贫血或艾滋病。我们要做的是让价格降低到 2000 美元，并且已经取得了惊人的进展。我们设定了一个十年计划，我们的目标是让非洲也可以负担得起这样的手术。

mRNA 也很值得研究。盖茨基金会目前正与三家 mRNA 公司——Curevac、Moderna 和 BioNTech 合作。BioNTech 是非常好的合作伙伴，我们在合作研发结核病和疟疾的 mRNA 疫苗。

在气候领域，核聚变能源对于优化能源供给结构非常重要。所以我们投资了 TerraPower 及其它四家核能公司。核聚变很有希望，因为它是一种 24 小时可靠、不受天气影响的绿色能源，如果它能够足够便宜，那将是一项非常重大的突破。同时，我们也在关注很多领域的排放问题，从钢铁、水泥、直播稻（direct seeded rice）等多方面助力减少排放。

过去五年，有关创新的工作，包括微软的工作、气候工作、

全球健康工作等方面的进展都比我预期得要快。我们正处于一个复杂的社会环境中，政府部门也需要考虑如何面对人工智能，并帮助人们过渡和适应新的形势。

周礼栋：技术创新的步伐正在加速，而我们必须把这种加速引向正确的方向，来解决真正的问题。非常感谢您来到中国，来到微软亚洲研究院与我们分享这些见解，谢谢！

相关链接：

[1] Butler Lampson、Charles Simonyi、Alan Kay，三人都是计算机科学领域的杰出人物，他们曾经在施乐帕克研究中心（Xerox PARC）一起工作，参与了许多创新性的项目，对个人计算机和网络的发展有重大影响。

周礼栋对话青海大学校长史元春：跨学科创新，与 AI 共成长

2023 年 5 月，微软亚洲研究院院长周礼栋博士来到了青海大学，与青海大学校长史元春及青海大学的师生们一起畅想了跨学科领域的创新之路。青海大学地处青藏高原中心，具有很多特色的学科专业，包括三江源生态学、藏医学、高原畜牧业等，这些专业要如何与人工智能、计算机技术深度结合？对此，周礼栋院长和史元春校长分享了他们各自的观点，并共同探讨了跨学科研究的核心驱动力、新技术背景下跨学科创新机遇等议题。

问：请问周礼栋院长，您对青海大学有什么印象？

周礼栋：此前史校长曾向我介绍过青海大学，那时我就对这所学校充满了无限憧憬。这次来到这里，我发现青海大学所取得的成就比史校长描绘的还要辉煌，尤其是三江源生态、藏医学、高原医学、高原畜牧业等专业让我印象深刻。这些专业不仅对青海省的发展有重要意义，也对整个中国乃至全人类的福祉有个很大贡献。同时，青海大学对中国西部地区的人才培养和经济建设也做出了卓越贡献。

另外，清华大学对青海大学的对口支援，也取得了显著的成效，帮助青海大学实现了每五年一次的飞跃。微软亚洲研究院也希望能够与青海大学建立紧密的合作关系，共同探索一些具有挑战性和创新性的研究，为青海省、为中国、为全球社会做出更多贡献。

问：作为微软亚洲研究院的老朋友，史元春校长从 1993 年在清华大学计算机科学与技术系任教，到 2022 年上任青海大学校长，和研究院始终保持着密切的联系。请问史校长您对微软亚洲研究院有怎样的印象？

史元春：在微软亚洲研究院成立之初，甚至还没有创立之前，清华大学就与微软公司开展了合作。微软亚洲研究院成立之后，我们不仅在物理位置上更接近了——与清华大学计算机系直线距离仅一公里多，而且我们的合作关系也更加紧密。在过去的 20 多年

里，微软亚洲研究院的“明日之星”计划为众多高校学子提供了实习机会，很多清华大学的学生都通过这个计划得到了锻炼和成长，例如，清华大学在人机交互领域顶级会议上发表的第一篇论文，就是我的一个博士生与微软亚洲研究院合作完成的。



微软亚洲研究院院长周礼栋（中）与青海大学校长史元春（右）在青海大学的对话，由微软亚洲研究院学术合作总监马歆（左）主持

微软亚洲研究院拥有国际化的视野和自由多元的学术氛围，致力于解决最具挑战性的问题，对于学生来说，这里是一个开阔眼界、修炼科研身心的地方。同时，在过去的 25 年里，微软亚洲研究院也从一个纯粹的计算科学研究机构，发展成了一个涉及多个交叉领域的研究机构。例如，使用 AI 技术解决甲骨文研究中的难题、进行 AI for Science 方面的研究等，展现了交叉学科研究的重要性。

上一任青海大学校长王光谦教授在与我交接时曾说“用 AI ‘扫荡’一切，让 AI 技术渗透到青海大学的每一个学科中。”我希望我们学校能有更多的学生，无论是计算机专业还是其他专业，都能积极参与到“明日之星”项目中，与微软亚洲研究院一起探索 AI 技术的无限可能。



青海大学校长史元春

问：两位在跨学科研究方面有哪些经验和案例？为什么跨学科合作如此重要？

史元春：先讲一下我对跨学科的认识。跨学科其实并不神秘，因为学科的设定是社会发展与科技发展的产物，它本身并不是一成不变的，在研究和发展的过程中，你会发现很多问题不能仅仅用本学科的方法来解决。尽管在某一学科的范围，我们都有自己的研究体系、教学计划，但是每个学科都经历过一个发展的过程，它是从无到有的，而跨学科就是这个发展过程的体现。

在青海大学，我们有很多专业和学院都是经过发展变化而产生的，其本身就具有交叉学科的属性。例如能源与电气工程学院，电力中的储能和调度会涉及到计算和控制技术，以及新能源研究等，这些就是计算机与新能源光伏领域的联合研究。

周礼栋：从研究创新的角度来看，很多最了不起的创新都是在现有学科体系之外产生的。微软亚洲研究院作为一个以研究创新为核心使命的机构，我们一直在寻求打破现有学科之间的界限，只有突破这些界限，我们的研究才能更具创新性。

另外，AI 将来必然会在很多方面为人类提供帮助，我非常钦佩史校长对青海大学的战略布局，包括启元实验室数据与计算平台（第四期），这是未来所有学科融合的基础条件。尤其是在计算机科学领域，在 AI 高速发展的阶段，我们看到了很多技术与其他专业领域的结合，这对于整个社会、经济以及人类的发展都将带来巨大的影响。

微软亚洲研究院现在也在研究 AI 与人和的关系，以及 AI 快速发展时，我们将可能面临怎样的未来挑战。这不仅仅是计算机科学本身应该要研究的课题，还涉及到社会学、心理学、哲学等社会科学，而这些领域本身也会因为人工智能和计算机技术的发展产生颠覆性的变化。我们希望通过跨界研究，真正了解那些对人类社会有重大影响的问题，并试图提供一定的解决方案。

在学生阶段，我们可能更专注于自己的研究领域，因为我们的目标可能只是发表一篇论文。但当我们把注意力转移到如可持续发展、医疗健康等对社会更有价值和影响的问题上时，就会发

现单一领域的技术是不够的。有了更高远的志向，不同领域的人就会为了同一个目标而努力，所有人就会共同协作解决问题，一起思考寻找打破领域界限的方法和解决这些重要问题的方案。

问：新技术的不断涌现将会给跨学科创新带来哪些机遇？

史元春：在开学第一课上，我给同学们介绍过，我们应该与 AI 共同成长。大模型、AI 让我们把原本在脑海中、在书本里的知识和智慧变成了可以使用的工具，支持我们进一步发展和创造。这类工具的出现，并不是最近几十年计算机出现后才有的，其实自古已有之。我们生活在一个充满机遇的时代，不论什么专业，我们都应该在学习和研究中充分利用这些可以承载人类基本知识的系统和工具。我们要带着问题、带着对自己能力发展的渴求、带着培养解决问题的能力的需求，开启新的学习和研究之旅。

过去已经有大量的研究成果是通过 AI 加持而产生的，例如刚才说的调度控制与计算的结合，还有清华大学研究的蛋白质大模型与我们农业育种大模型的结合，帮助我们改进了马铃薯、油菜、蚕豆等蔬菜的育种技术。今天的许多学科的范式未来必然会用到 AI 技术，从学校角度来说，青海大学加强了算力和智能算力平台的建设，以及团队和支撑能力的建设。我希望通过与微软亚洲研究院的深入合作，可以充分利用国内外资源开展深度交叉研究。总之，不要把 AI 当成玩具，而是要尽快把它作为自己学科研究中的工具。

周礼栋：如今我们所身处的这个时代充满了创新的机遇。AI 将会颠覆很多领域，但它的颠覆并不是从领域外进行的，而是需要 AI 和每个领域紧密结合才能够发现这个领域的新未来。

青海大学有着独特的优势，拥有很多特色专业，有很多重要的问题就摆在面前。同学们要努力在各自的领域扎根，找到那些你们认为重要的、并抱有最大热情的问题，也可以加入微软亚洲研究院与我们共同探索人工智能和其他专业领域的结合，重新定义领域的未来。这是一个相互学习的过程，也是一个艰难而漫长的过程，因为不同领域的知识语言系统是不同的。但这也是一个必然的过程，通过将 AI 与领域知识紧密结合，我相信未来可期。



微软亚洲研究院院长周礼栋

问：人工智能技术该如何与生态学研究更好地结合，为碳达峰、碳中和做出更大的贡献？

周礼栋：生态环境和可持续发展是当今社会面临的最重要问题之一。微软亚洲研究院和清华大学做了很多相关的探索，比如我们和清华大学地球系统科学系合作开发了全球首个近实时负碳数据库，并且提出了一种机器学习 + 运筹优化的解决方案，帮助煤电实现低碳转型；我们还与清华大学地球环境学院共同研发了 DeepRSM 大气响应模型，能够更精细地刻画空气污染物浓度，从而帮助快速找到最佳的减排方案等等。

我们也非常希望与青海大学就三江源生态问题展开联合研究。这是一个复杂的问题，随着 AI 的进一步发展，它对生态领域的支持作用会更强，同时我们面临的挑战也会越来越多。我们欢迎青海大学的老师和学生与我们共同寻找利用 AI 解决生态问题的方法，找到新的解决方案。

史元春：青海大学研究生生态不只是单纯的生物多样性和环境保护研究。青藏高原是所有高原人民的高原，生产、生活和生态都在这里同时发生，比如碳达峰、碳中和、碳排放的问题，包括我们的载畜量、牦牛管理等。这里不仅有养殖业，也有对环境和双碳目标的影响，还有要如何提高居住在高原的人民的生活水平等。这是一个很难的问题，难点在于如何利用计算机技术，在约束条件较多的情况下去做多目标优化。今天计算机和 AI 技术为这个问题建立了大数据基础，而 AI 如何来加持这项研究，需要我们进一步地探索。

周礼栋对话 Chris Bishop：转变思维方式，重塑科研文化

2022 年，微软研究院成立了科学智能中心（AI4Science），专注于将科学发现的第五范式变为现实。2023 年，周礼栋博士与微软技术院士、微软研究院科学智能中心负责人 Chris Bishop 进行了一次深入的交流。Chris Bishop 分享了他对科学智能中心的期待，以及 AI for Science 实践中的机遇和挑战。他们还探讨了在科学研究新范式崭露头角时，研究机构应如何转变思维，重塑科研文化，以促进跨学科合作，从而让科学研究对世界产生更积极的影响。

周礼栋：非常高兴能与 Chris 进行这次对话。微软研究院科学智能中心（AI4Science）已经成立近一年了，团队的发展是否达到了你的期望？科学智能中心有怎样的发展规划？

Chris Bishop：成立科学智能中心是微软研究院的一次“实验”，我们还处在起步阶段，发展道路上必然会充满各种挑战，我们还有很多事情要做。但总体来说，科学智能中心的发展非常令人满意，我百分之百确定它的成立是微软做出的一个正确的决策。

我们都知道，科学的发展并不是在所有领域和所有时代都是匀速的，在某些领域在某些时期会出现爆发式的进步。毫无疑问，未来五年甚至更长的时间里，深度学习将会推动自然科学取得非常迅猛的发展。这也是我们开展这个有趣“实验”的原因之一——建立一个全球化的团队，专注于 AI for Science，我们的研究不受地域限制，这与微软研究院的传统组织方式有所不同。

AI for Science 的研究对微软至关重要，也是公司的战略布局。目前为止，科学智能中心的进展非常顺利，我们正以每小时 500 英里的速度向前奔跑，因为科学创新也是“唯快不破”。



微软亚洲研究院院长周礼栋（右）与微软技术院士、微软研究院科学智能中心负责人 Chris Bishop（左）

周礼栋：微软首席科学官 Eric Horvitz 曾说过，微软研究院发展的关键原则是开放、合作和自由的研究氛围。多年来，微软研究院的研究员一直在这样的环境中工作。但近年来，使命驱动的研究成为了一个新的趋势，它与传统的开放、自由的研究模式有所不同。你认为如何在自由的研究和使命驱动的研究之间更好地取得平衡？

Chris Bishop: 研究机构有很多不同的运营模式，不同研究机构多年的实践也证明了没有一种模式是绝对正确的，不同模式在不同的时期和不同环境下都可能取得成功。

1997 年，微软开设了美国以外的第一个研究院——微软剑桥研究院。那时剑桥研究院只有三名员工，他们都来到了当时我所在的艾萨克·牛顿研究所（Isaac Newton Institute），邀请我加入他们。我大概只思考了几秒钟，就决定加入这个新兴的研究机构。

我之所以提到微软剑桥研究院，是因为它有一位杰出的创始人 Roger Needham。Roger Needham 从本科开始就一直在剑桥大学工作和生活。他按照自己熟悉的类似于大学的方式来运营微软剑桥研究院，比如相同的组织结构和激励机制。我认为这是一个正确的模式。如果是我，我也会采取相同的方法。

如今我们正处在人工智能、机器学习的时代，微软在这些领域有着领先的研究和产品，大家希望来微软工作是很自然的事情。但在 26 年前，微软要怎样吸引一个对机器学习感兴趣的理论物理学家加入一家软件公司呢？这样一种更加学术化的机制是吸引优秀科研人才加入微软的重要方法。过去几十年中，微软研究院一直稳步发展，保持增长，也建立了新的研究院。随后，我们进入了一个发展相对平缓的时期。

微软是一家坚信技术和创新都至关重要的公司，这也是微软研究院存在的根本原因。只有创新才能生存，而研究就是创新的源泉。因此，微软必须开展研究工作。问题在于，是需要一个独立的研究组织，还是把研究嵌入到产品部门中？我认为应该两者兼顾，事实上，微软也是这么做的。

两者之间的区别在于，微软研究院应该专注于那些只有我们才能做到的事情，也就是更具颠覆性和变革性的研究，尽管这些研究可能伴随着更高的风险，甚至在大多数情况下不能取得预期的结果。我们肩负重要的使命，除了与产品团队紧密合作，将一些技术交付给产品团队之外，还要作为一个独立的组织，开展更具雄心的工作，确保微软始终在创新的前沿。

周礼栋: 如你所说，在微软研究院，我们做的都是非常前沿的工作，具有极大的风险。你认为如何改变文化，转变思维、心态或者工作流程，来支持微软研究院的变革，完成科学智能中心的使命？

Chris Bishop: 首先，明确一个方向非常重要。我们已经定义了科学智能中心的使命，即通过深入的研究，对一些关乎人类命运的重大课题，包括气候变化、药物发现等，取得突破性的进展。想要对全世界产生积极的影响，在这些问题上有所突破，必须进行规模化研究，而微软提供了实现规模化研究和达到这一目标的理想平台。我们要将其打造成一个有前景的业务，并确保其可持续发展。

我们之所以如此兴奋，是因为我们可以通过积极的方式推动世界的发展。要想取得成功，我认为有三个要素至关重要：首先

研究必须具备深度，这样才能在产品之外有所作为；其次，要借助微软的平台实现规模化，让研究更具可持续性；最重要的是，我们的研究必须对现实世界产生积极的影响，这也是我们最关心的。这些因素构成了我们当前工作的核心内容。



周礼栋: 合作在当前的环境下非常关键。我们要如何鼓励大家进行合作？特别是在跨地区和交叉学科研究的背景下，如何通过合作取得更大的成就？

Chris Bishop: 这涉及到自下而上和自上而下的研究模式。纵观微软研究院的历史，尤其是我所熟悉的剑桥研究院，所有研究员都有充分的自由，可以专注于自己感兴趣的问题，这是一种自下而上的研究模式，促成了百花齐放的成果。

现在很多事情都在发生变化。例如，机器学习初期的神经网络可能只有三四百个节点，而现在的规模已经超出了我们的想象。要取得世界级的研究成果，我们需要进行跨学科、跨团队的全球合作，这需要出色的项目管理。

因此，我们也需要融入自上而下的模式，但这并不意味着我们要告诉研究员应该从事哪些研究。研究员们都非常聪明，对自己的领域了解更多，管理层的作用更像是倾听者，并提出自己的期望。

这也涉及到如何激励人们更好地开展团队合作。我们鼓励大家更大胆和更具雄心，冒险就意味着有可能失败。但是如果一个团队团结一致朝着合理的目标努力，即使最终因为无法控制的因素而失败，我认为这是可以理解的，团队也应该获得奖励。而在微软，这也不是难事，因为我们的三条相互关联的奖励原则中，只有一条是关于个人技术贡献的，另外两条都与他人协作有关，我们会奖励那些在团队中合作的人。

当然，合作也意味着要放弃一些自主性、自由和能动性，但有时我们确实要做出妥协，以获得更大的回报。有时候看似放弃一座金山，实际上可能会获得更宝贵的财富。换个角度看，我们不必独自发表大量论文，而是可以加入一个更具影响力的团队，成为一篇颠覆性的论文中的众多作者之一，与他人一起改变世界。

在协作方面，科学智能中心就面临跨时区工作的挑战。我们

的全球团队分布在三个不同时区，这可能给团队成员带来很多困难，我们需要在实践中找到一种更明确和更本地化的工作流程。我认为，拥有优秀的项目管理专业知识可以帮助我们梳理工作流程，以更异步的方式思考，我们会逐渐提高效率。在微软研究院 30 多年的历史中，这也是前所未有的一次尝试。

周礼栋：我们注意到当前研究领域的一个显著变化是，我们的研究已经超越了计算机科学的范畴，跨学科的研究变得越来越重要。你一直在推动不同领域和学科之间的交流与合作，请分享一下你在跨领域和跨学科工作的经验，以及我们要如何弥合跨学科之间的鸿沟？

Chris Bishop：这是一个我们仍在不断探索的问题。在为跨学科研究而努力的过程中，我们领导团队意识到，要将人才聚集在一起，需要转变思维方式，从强调不同学科形式化结构上的区别，转向以使命和任务驱动。如果我们的目标是在现实世界中应用人工智

能，那么我们不仅需要机器学习，还需要工程学知识。同时，我们还必须考虑人工智能对人类的影响，这就需要社会科学的参与。

我们正在努力将不同的学科融合在一起，但必须承认，这些学科之间存在着很多差异。同时我们还要面临不同的语言、文化，跨区域协作等多方面的挑战。这些问题都很棘手，有时候需要放弃一些东西，虽然困难，但却是值得的。因为只有通过真正的跨学科合作，全球化的运营、吸纳不同学科、不同地区的顶级人才，才能做出真正能够改变世界的研究成果。

我没有所有问题的答案，但我们有一个明确的愿景，我们也清楚地知道只有通过团队协作，我们才能取得成功。我们已经踏上了这条困难重重的道路，并将坚定不移地砥砺前行，寻找解决问题、改变世界的途径。

Ashley Llorens 对话薛澜： 推动人工智能正向发展，跨领域与国际合作至关重要

2023 年 11 月 10 日，微软亚洲研究院与清华大学苏世民书院联合举办了一场主题为“促进人工智能的合作与学术交流”的座谈会。会上，苏世民书院院长薛澜教授和微软研究院副总裁、微软杰出首席科学家、微软研究院全球研究合作负责人 Ashley Llorens 博士就人工智能未来发展所需的合作展开了深入讨论，并分享了他们各自对跨领域与国际合作、人才培养等问题的见解。本次对话由微软亚洲研究院资深首席研究员谢幸博士主持。



微软研究院副总裁、微软杰出首席科学家、微软研究院全球研究合作负责人 Ashley Llorens（中）与苏世民书院院长薛澜（右）的对话，由微软亚洲研究院资深首席研究员谢幸（左）主持。

谢幸：人工智能技术和相关工具的日益普及，引发了人们对其可能带来的机遇和潜在风险的关注。两位对人工智能发展的治理有什么看法？

薛澜：作为一名政策分析者而非计算机科学家，我对人工智能技术的细节了解不多，但我一直在密切关注新技术可能给社会带来的影响。我曾参与中国在 2017 年发布的《新一代人工智能发展规划》起草过程中的讨论，并提到了发展人工智能需要关注潜在的风险。大家可以看到，这份规划中明确地强调了要关注人工智能发展可能引发的风险和相关的治理需求。后来，我被任命为新一代人工智能治理专业委员会的主任，并一直从事人工智能治理相关的工作。

许多人都担忧人工智能可能带来的潜在风险，但也有大量技术专家担心，政府的过度监管可能抑制技术的发展。如何在两者之间取得平衡，是我们当前面临的一大挑战。对此，我们一直倡导将敏捷治理作为人工智能的治理方法。

由于人工智能的发展给人类社会带来了诸多益处，所以推动其创新发展是毋庸置疑的。但同时，未知的风险也是巨大的，我们不能放任其野蛮发展，关键在于我们的治理如何适应并跟上这一新技术的发展脚步。因此，我们需要审慎地观察人工智能的发展，在必要时采取措施引导规制，同时避免过度干预阻碍其发展。

Ashley Llorens：我在人工智能领域有 20 年的工作经验，亲眼见证了机器学习技术和机器人应用将我们带入了一个激动人心的时代。我们正在经历技术史上的一个历史性时刻，一个关键的转折点。

回顾几十年来计算技术的发展，人们开发了各种应用程序来完成特定任务，现在我们已经创造了非常复杂的方式来显式地编程，让机器执行任务，例如使用 Windows 系统与各种应用程序和软件进行交互，这些程序将点击的动作映射为用户的实际意图，比如打开 Word 文档等。

随着机器学习技术的崛起，系统可以通过数据和经验实现自我编程、自动创建映射、自主完成任务。实际上，经过学习训练的程序就是一个高度参数化的数学函数，但在大语言模型和多模态模型出现之前，一个训练过的程序只能完成一项任务，比如一个计算机视觉算法，它接收像素输入并输出分析结果，描述图像中包含的对象或预测视频序列中要执行的动作。但现在，自监督机器学习的过程可以创建执行多个任务的程序，而且还可以使用自然语言对它进行再编程。ChatGPT 带来的改变是革命性的。我们可以使用开发人员提供的基础模型，以前所未有地方式使用这些工具，或者用开发者未曾想到的方式进行创新。

回到人工智能的机遇和风险这一问题上。用户可以更具创新性地来使用这些技术工具，但这也带来了风险，比如，用户想用工具来做一些不好的事，或者因为对模型能力的片面理解导致模型产生了意外的结果，尤其是在医疗等关键领域，有些意外结果将是危及生命的。

因此，技术与人类需要共同进化，在利用技术优势的同时，也要基于负责任的人工智能原则来进行创新。微软推出的大语言模型产品都经过了严格的测试，就是这种共同进化的例证。

谢幸：我想大多数人都会认同这是一个技术变革的时代，也意识到了机遇与挑战的并存。人工智能的快速发展，特别是大语言模型的兴起，既彰显了技术的进步，也引发了很多担忧。两位对人工智能的未来有什么看法？是乐观更多一些，还是会有些悲观？

薛澜：如果没有人关心人工智能可能带来的风险，那才是真正的令人悲观。相反，正因为现在有很多人关注这个问题，才让我对人工智能的未来感到乐观。



苏世民书院院长薛澜

以核技术为例，在其发展初期，科学家们也为核弹的发展展开过争论，当人们看到核弹的破坏力之后，便开始倡导停止这项技术，并形成了一整套控制核武器扩散的机制。由此可见，人类社会是有能力控制技术的应用，限制技术向负面方向发展的。

同样，在生物技术领域，基因编辑等前沿技术能够完成很多超乎想象的事情，但也伴随着伦理道德方面的担忧。而现在医学界和生命科学界已经尝试提出一套伦理方针来指导研究。人工智能科学家和政府机构也在努力探索治理机制，来为技术提供防护栏。因此，我对人工智能的未来还是持乐观态度的。

Ashley Llorens：我同意薛教授的观点。我从事人工智能和机器学习相关工作已经有很长时间了，如果 10 年前我说自己是做机器学习的，很多人会感到困惑。但现在，人们对人工智能的发展有了更多的了解，而且计算机领域之外的人每天也在利用这项技术解决实际问题。我们也看到在学术界，很多人在推动人工智能技术的发展。人类的一个优势是擅长使用工具，从某种角度来说，人工智能作为一种工具，是我们与物理世界交互的延伸，也是我们认知能力的放大器。

同时，我们也需要保持警惕。乐观、怀疑、悲观，这些不同的态度都是正常的。当我们从整个社会的角度来考虑资源投入时，要把这些态度都纳入考量，关键在于投入的分配比例。尽管一颗小行星撞击地球的可能性非常小，但美国国家航空航天局 (NASA) 还是投入了一些资金，通过小行星重定向项目来降低相关的风险。我们需要有一个多元的投入组合来反映我们对先进技术的不同态度，并且要确保我们的决策不会阻碍技术的正向发展。

谢幸：清华大学和微软研究院都是高度国际化的组织，并且有丰富的国际合作经验，你们如何看待国际合作在应对人工智能风险和挑战方面的作用？以及如何让合作更为有效？

Ashley Llorens：国际合作在人工智能领域至关重要。我们需要找到最佳的合作方式，以实现更大的成功和积极成果，同时避免负面影响。微软一直致力于推动人工智能技术的正向应用，例如我们发起的“气候研究倡议 (Microsoft Climate Research Initiative, MCRI)”，就是旨在利用人工智能技术应对气候变化。在内部，我们与可持续发展研究领域的同事合作，确定了人工智能和计算技术可以帮助应对气候变化挑战的三个研究方向：碳核算、碳减排和气候韧性。在外部，我们与高校合作共同探索如何结合新的遥感技术和人工智能进行碳核算，其中就有来自清华大学的研究员。这是我们通过合作推动技术正向应用的一个例子。

薛澜：从全球视角看，近年来科学家之间的国际合作日益增加。在研究层面，合作尤为重要，我们的研究发现，在科技文献发表方面美国和中国是合作最多的国家。在机构合作方面，就像 Ashley 提到的，清华大学的科学家加入到微软资助的研究项目中，还有很多清华大学的毕业生在微软工作，微软也和清华大学在很多项目上开展了合作。在国家层面，我们要避免极端竞争的情况出现，给科技合作提供基本保障。

谢幸：除了国际合作，跨学科合作也是人工智能发展的重要维度。越来越多的研究人员意识到，要推动人工智能的进步，必须与社会科学等其他学科相结合，以便更深入地理解和应对人工智能对社会的影响。请两位谈谈你们对跨学科合作在人工智能发展中的作用和意义的看法。

薛澜：我们今天的座谈会就是一次跨学科对话与合作的例证。在人工智能治理方面，我们需要技术专家、哲学家、政策制定者和管理人员的共同参与，跨学科合作对人工智能的发展非常关键。

AI 对公共管理和社会科学研究的影响正受到越来越多的关注。很多人认为，AI 带来了巨大的范式转变，新技术改变了社会科学研究的格局。因此，我们需要加强不同学科间的合作，清华大学也成立了社会科学和计算机科学的跨学科研究机构。

Ashley Llorens：跨学科合作是解决人工智能领域中核心问题的关键。作为人工智能的科研人员，我认为探索通用智能的本质会带来很多有趣的技术挑战，其中最有趣的科学问题是关于人类的问题。还有一些更广泛的社会问题，比如社会是否应该与这些技术共同进化。这些问题非常有意義，也是非常重要的问题。



微软研究院副总裁、微软杰出首席科学家、
微软研究院全球研究合作负责人 Ashley Llorens

微软研究院很早就意识到了这一点，我们汇集了很多社会科学家，包括计算社会科学家、经济学家、人类学家等各领域的专家。我们还需要更多来自不同领域和不同背景知识的人才，来共同构建一个跨学科合作的社区。

谢幸：鉴于跨学科合作的重要性，我们需要培养更多的跨学科人才。然而，目前这方面的人才还比较匮乏。我们应该如何为未来的人工智能培养跨学科人才，以应对未来的变化和挑战？

薛澜：其实这是一件很容易的事，可以先在苏世民书院学习，然后加入微软公司工作（笑）。

Ashley Llorens：我们必须处理好人工智能对教育的影响，并充分利用其为教育带来的机遇。从小培养孩子的算法思维和系统性思维，比以往任何时候都更重要，技术可以帮助我们实现这一目标。

现在，机会就在眼前，我们需要加大这方面的投入和支持。

同时，我们也需要在学校中规范地使用人工智能技术，比如在论文撰写中合理运用 AI。我们不应过度依赖工具，而忽视必要技能的培养，当然我们也不能忽视对新兴技术的使用。

薛澜：未来，人工智能有望帮助消除不同领域之间的障碍。人们常常有一种固有印象，认为学习科学和数学知识之后，再学习人文和社会科学是件很容易的事，但反过来却比较困难。现在人工智能也许可以改变这种领域偏见。人们可以平等地使用各种工具来学习科学、数学、工程和社会科学等，或许这是未来使用人工智能工具就能实现的。

Ashley Llorens：我完全同意。我相信，未来人们可以更容易地享受到技术进步带来的好处，使用自然语言编程系统将成为一项基本技能。尽管这还存在挑战，但微软研究院正朝着这个目标努力，到那时不仅计算机科学家可以用编程语言编写代码，非计算机科学家也能用自然语言为代码库做出贡献。未来 5 到 10 年，我们对技术人员的定义可能就会完全不同。

洪小文对话图灵奖获得者 Raj Reddy: 人工智能将改变我们学习方式

2023 年 11 月 10 日, 图灵奖获得者、人工智能领域先驱、卡耐基梅隆大学教授 Raj Reddy 访问微软亚洲研究院并作分享。微软亚洲研究院院长周礼栋博士对 Raj Reddy 教授多年来的支持表示感谢, 并送出一份特制的纪念相册, 其中记录了 Raj Reddy 教授与微软亚洲研究院 25 年以来的点点滴滴。在随后的对话环节中, 微软全球资深副总裁、微软亚洲研究院前院长洪小文博士与 Raj Reddy 教授对技术的发展和未来进行了深度探讨。Raj Reddy 教授分享了自己对当前快速发展的人工智能的见解, 预测了未来人工智能的研究趋势及社会影响, 并与洪小文博士共同探讨了如何拥抱和引领下一个时代。



图灵奖获得者、卡耐基梅隆大学教授 Raj Reddy (左)
与微软全球资深副总裁洪小文 (右) 进行对话

洪小文: 您是人工智能领域的先驱之一, 已在领域内耕耘了 60 年, 请分享一下您研究人工智能 (AI) 的经历。

Raj Reddy: ChatGPT 与当年的 n-gram 模型有很紧密的关系, 甚至 ChatGPT 的底层仍然是 n-gram 模型。60 年前——也就是 1963 年——DARPA (美国国防部高级研究计划局) 认为 AI 非常重要, 于是资助了麻省理工学院、卡内基梅隆大学和斯坦福大学进行持续研究。当时我还是斯坦福大学的一名研究生, 而这就是我研究人工智能的开始。但当时的研究并没有那么严肃, 学者们研究 AI 更多地是出于个人兴趣, 例如, 他们会创建一个 AI 来帮助自己玩象棋游戏。之后, 我们获得了更多的资金。人工智能不仅仅是用来创造智力游戏, 还有听说读写等很多人类擅长的事情需要人工智能去尝试。我刚开始接触计算机的时候, 电脑有一个房间那么大, 内存很小, 且频率很低。现在, 个人电子设备已经可以放在口袋里了, 内存大了很多, 频率也变快了。这是一个非常有趣的经历。

洪小文: 现在的大模型可以完成很多通用任务, 它可以开发很多工具来帮助人类社会。您对此有何看法?

Raj Reddy: 在微软, 你们有很好的机会去接触 GPT-4, 微软亚洲研究院也正在针对特定领域 (例如医疗、生物等) 训练专门的大模型。这个领域有很多探索的空间。我鼓励大家多花一些时间去探索自己感兴趣的事, 并将自己的才华运用到正确的方向上。

有几个问题我一直在思考, 但始终没有想出准确的答案。

第一个问题是, 当大模型拥有大部分知识的时候, 人类为什么还要去学习? 知识有两类, 一类是事实性知识, 例如数字、时间等, 另一类是技能性知识, 例如弹钢琴。人工智能可以帮助我们获取事实性知识, 也可以指导我们如何练习技能性知识, 但我们必须亲自去实践, 才能学会这些技能。而且, 人工智能并不总是正确的, 它会产生错误或虚构的内容。因此, 我们需要具备判断能力, 需要仔细地、批判性地阅读人工智能生成的信息, 再决定是否采信它的回答。只有这样, 我们才能真正利用好人工智能。提出问题、不断挑战自己所看到的每一个信息, 这对于人类而言是非常重要的。未来的世界将充斥着虚假信息, 技术可能被恶意利用。所以, 我们必须培养自己的思考能力, 不要盲目相信我们看到或听到的一切, 要敢于提问, 要有自己的判断和观点。

第二个问题是, 在未来我们应该专注于学习什么? 我认为应该学习技能, 而非学习知识, 因为所有的知识都将可以通过人工智能轻松获取。现在的很多考试都是不合理的, 因为他们只考察我们对于事实的记忆, 而不考虑我们是否能够随时查找到这些事实。我认为, 人工智能、深度学习等技术不仅会改变我们所知晓的知识, 也会改变我们学习的方式。整个教育体系都应当做出调整, 侧重于让学生学习技能本身。而且, 今后的教育可能也不需要让学生们都坐在教室里了, 大家可以远程上课。



图灵奖获得者、卡耐基梅隆大学教授 Raj Reddy

洪小文：您认为，在未来，年轻学生需要关注的重要的研究问题是什么？

Raj Reddy：第一个问题是语音。目前，世界上有大约 10 亿人不具备读写能力，只能通过说话来交流。他们无法在互联网世界发挥任何作用，被迫降级为“二等公民”。我们或许可以利用基于语音的对话系统，设计一个适合这类人使用的界面，让他们能够与人工智能进行对话，帮助他们获取互联网信息和服务，例如在线购物、旅行等。读写能力的差距是可以被解决的，技术可以让世界上的每个人都能参与到互联网的活动中。

现在，我想来谈谈语言。世界上有数百种语言。我有一个愿望——借助 AI 技术提供的翻译服务，可以让我们读任何书、看任何电影、和来自任何地方的任何人进行交谈，而不受限于语言的不同。

最后是小语种和方言的翻译问题。许多小语种和方言缺乏足够的语料库。AI 如何利用有限的训练资源学习小语种，并提供小语种的翻译服务？消除语言代沟会为人类社会带来很大帮助。

洪小文：很多人觉得 AI 过于强大，认为我们应该对人工智能进行监管。您对此有何看法？



微软全球资深副总裁洪小文

Raj Reddy：这是一个非常重要的问题，我对此保持积极乐观的态度。我认为，人工智能或许确实会给社会带来一些麻烦，但不会对人类产生毁灭性影响。就像每年都有很多人死于交通事故，但我们并没有停止使用汽车。

人工智能也可能拥有偏见。因为人类本就有偏见，而人工智能又是从人类生产的海量数据中学习的。这肯定会引发一些担忧，但并不意味着应该停止对人工智能的研究。当汽车刚出现的时候，人们并不需要考驾照，在很多交通事故发生之后，人们才意识到驾驶汽车之前要经过培训。偏见问题固然存在，但这不是人工智能的问题，而是一个社会问题，我们应该让使用这些大数据的人去解决。

现场问答



现场观众提问

Q1：有些人认为通用人工智能会导致学生在作业中作弊，有些人则认为会对学生有很大帮助。我想知道，为什么这个讨论存在很大分歧？如何有效地将人工智能应用到教育中？

Raj Reddy：我一直告诉人们，在未来，“抄袭”应当是可以被接受的。与其给人添乱，不如鼓励人们在最短时间内找到答案、解决问题。我希望你们都具备这种能力。

洪小文：在 GPT 可以帮助大家写作的时代，一些人认为我们没必要再学习英语了。这是不对的，英语反而变得更重要了。第一，面对两篇 ChatGPT 生成的文章，你需要对文本有足够的理解，才能判断孰优孰劣。第二，为了得到更好的输出结果，你需要对 ChatGPT 给出准确、高质量的输入。你需要用语言清晰地描述自己的诉求——如果你都不知道如何描述你的问题，别人又怎么帮助你呢？大部分大语言模型的训练内容都是英语，因此大模型在英语内容上的表现会比其他语言更好。在“理解”这一层面，能够区分 A 和 B，能够判断 A 比 B 好，需要很多知识和技能。

Q2：对于那些想要跨学科从事人工智能研究的学者，您有什么好的建议吗？

Raj Reddy：我认为有两件事需要做：一是学习编程思维，二是掌握基本的阅读和写作能力。现在的教育体制太过僵化，每个人都必须上大学，在固定的时间段内修读固定的课程。我们需要重新设计教育系统，让每个人能够根据自己的兴趣灵活选择不同课程。这样，大家可以根据自己所学的知识去做更多有意义的事情。

AI 与社会、科学的协同：跨学科研究的潜力与挑战

2023 年 11 月 14 日，微软亚洲研究院与东京大学联合举办了以“人工智能协同：社会与科学 (AI Synergy: Society and Science)”为主题的 2023 年人工智能论坛。在圆桌讨论环节，东京大学新一代智能科学研究中心负责人、信息科学系研究部智能机械信息技术学教授 Yasuo Kuniyoshi（国吉康夫）、东京大学神经智能国际研究机构特聘教授 Yukie Nagai（长井志江），以及微软亚洲研究院副院长邱锺力、微软亚洲研究院全球研究合伙人韦福如、微软亚洲研究院资深首席研究员谢幸、微软亚洲研究院资深首席研究员边江，共同探讨了人工智能对社会和科学的重大影响，以及人工智能在解决复杂社会挑战和促进科学发现方面的潜力和作用。

谢幸：人工智能不仅是一项技术奇迹，也是助力我们改变世界的催化剂。从教育、医疗到娱乐，人工智能与不同领域的融合都给人类社会带来了巨大的变革，为我们带来了前所未有的机遇和挑战。那么，先请各位畅想一下人工智能的未来发展，它将带来哪些影响？



微软亚洲研究院资深首席研究员谢幸

国吉康夫：上个月我在《新科学家》(New Scientist) 杂志上看到了一篇文章，介绍了当 ChatGPT 在回答用祖鲁语提出的恐怖袭击问题时，竟然也给出了一些建议。如今，几乎每周都有关于如何欺骗大语言模型以及如何应对这些欺骗的论文发布。AI 正在不断对齐人类的价值观，但我认为这还不是最终的解决方案。

我一直从事有关具身认知 (Embodied Cognition) 产生和发展方面的研究，专注于构建能够模拟婴儿的具身模型和大脑模型，希望通过感知和运动的交互，让模型能够自主发展。目前，我们正在努力让模型拥有初级的道德意识。

这看起来似乎还很遥远，但是发展心理学的研究表明，6 个月大的婴儿就已经具备道德判断和惩罚坏人的能力。这意味着在

掌握语言之前，人类已经具有道德意识，这或许有助于我们在一开始就让具身交互的人工智能系统具备道德意识。这一研究涉及多个跨学科领域，包括人文学科、心理学、神经科学、人工智能、机器人等，最终可能发展成为一门关于人类的建构性科学。

在人工智能快速发展的背景下，对人性的探讨将成为一个重要问题。每个人都应该思考人类存在的意义，以及如何让人工智能与人类价值观对齐。为了解决这些问题，我们需要广泛的合作和持续的研究，也许我们能够创造出真正理解人性与人类共同发展的人工智能系统。

邱锺力：我们与计算机交互的方式正在发生变化。过去，我们主要是依靠键盘输入信息，而未来，我们的动作、面部表情、呼吸、心跳，甚至脑电波信号，都有可能被计算机感知。通过对自身情绪、健康状况、幸福感以及周围环境的感知，我们可以相应地调整环境，让自己更加舒适。比如，当我们呼吸急促，心跳加快时，系统可以感知到我们可能正面临压力，然后通过调节环境温度或播放舒缓的音乐来帮助我们放松。再比如，我们可以通过某些信号来判断一个人是否有抑郁症或者是否有早期症状，那么系统就可以建议用户尽早就医进行检测或诊断。

当了解到用户更多的详细信息时，我们就可以为用户提供更个性化和有效的医疗服务。人工智能将给未来的医疗带来革命性的变化。

当然，我们在使用人工智能时仍需谨慎，因为人工智能也会犯错，人工智能可以提供关键信息供医生参考，关键决策还需要由人类医生做出。



微软亚洲研究院副院长邱锺力

长井志江：人工智能在工业领域取得了显著进展，为人们的工作和医疗诊断提供了有效的辅助。然而，这些进展主要集中在特定的领域，我们还需要从整个社会的角度思考，如何让人工智能真正改善我们的日常生活，尤其是我所关注的人工智能技术和机器人的社交互动方面。

发展心理学研究显示，婴儿的许多感觉和运动经验并不是建立在语言基础上的，在他们还不会说话的时候，就能够学习大量知识。那么，这些经验知识是否能够帮助我们丰富人工智能系统的交互模式？毕竟，目前的语言模型还无法处理多种模态的交互。语言是人脑的一种抽象表达，而不是基于感觉运动经验的。在社交中，我们会接收到丰富的感觉信号，除了符号语言，还有视觉、听觉、触觉等多种感官刺激，以及身体内部的一些状态，比如，情感就与我们身体内部的表征密切相关。

对于大语言模型，我们还需要思考能从哪些方面来完善它，以及如何将语言模型与我们的经验建立联系，让人工智能技术不仅仅应用于特定的医疗或工业领域，而是让人工智能技术与人们的日常生活紧密连接，并从中受益。



东京大学神经智能国际研究机构特聘教授长井志江

边江：微软亚洲研究院与一些产业界的合作伙伴建立了紧密的合作关系，涵盖金融、物流、能源、汽车和医疗等多个领域。我们发现，大语言基础模型的蓬勃发展，使我们能够模拟更加复杂的真实世界。



微软亚洲研究院资深首席研究员边江

例如，通过与金融企业合作，我们利用基础模型来模拟金融市场，通过模拟特定的金融场景，甚至是一些极端情况，来测试我们的技术能否应对这些挑战。构建逼真的数字世界，为不同产业领域提供支持，推动产业发展，也是人工智能发展的一个方向。

谢幸：人工智能正在与社会科学、自然科学、健康医疗、脑科学等领域交叉融合。那么，各位专家认为人工智能将如何影响和改变计算机科学之外领域的研究和实践？

长井志江：人工智能与其他学科融合的一个成功案例是借助人工智能技术处理医学、人文科学等领域的大数据。人工智能拥有强大的计算能力，能够利用的知识量远超人类，因此可以更快速和更准确地分析数据。同时，我们也需要探索人工智能的新应用，改变人类研究或神经科学研究的方式。

邱锂力：我们的团队与上海的几家医疗机构有着密切的合作，积累了一些人工智能在医疗领域的应用经验。我们发现，医生在监测癫痫患者是否发病时需要全天紧盯屏幕。为此，我们开发了一种基于脑电信号（EEG）的癫痫自动检测系统，可以大幅减轻医生的负担，让他们能够专注于更重要的事情。这个系统还可以用于培训新一代的医护人员。

此外，人工智能还可以自动生成医学报告。现在我们需要专家来解读医学图像、X射线或CT等资料，但可以解读的专业医生数量是有限的。通过人工智能，我们可以让解读过程自动化，甚至还可以得到更准确的结果，提供更全面的诊断报告。我们希望未来人工智能可以成为医生的得力助手。

边江：在我们与产业界的跨学科合作中，会涉及到知识整合、知识生成等多个维度，这意味着每个行业都有大量特定领域的知识。例如，我们与物流企业合作，利用物流领域的知识库构建了一个能够解决物流问题的基础模型。这其中的挑战在于如何将领域知识库整合到当前的基础模型中。

另一点我想强调的是知识生成，就是在特定领域现有知识库的基础上，利用基础模型生成新的知识。我们正在尝试利用基础模型来实现工业企业的研发过程自动化。借助基础模型强大的推理能力，我们可以创造出特定领域的新知识。相信基础模型将在工业领域发挥巨大的作用，成为其发展的基石。我们很乐意与各领域的企业和研究人员共同开展跨学科、跨领域的合作。

谢幸：人工智能、大语言模型已经改变了许多行业的格局，自然语言处理、计算机视觉、系统等领域都受到了影响。在大语言模型时代，人工智能的研究是否会有一些不同的方法和重点？

韦福如：我们已经看到人工智能技术的范式正在发生转变，所以我们确实需要关注研究模式和组织形式的变革。人工智能不同研究领域之间的界限正在变得模糊，各种技术和模态也开始融合。因此，在这个新的时代，我们需要以更通用和更本质的方式来思

考问题，这意味着我们不能仅满足于针对特定任务 5%、10% 改进的增量式研究，而是需要思考能带来十倍，甚至百倍提升的研究方法，并重新定义研究主题，从更广泛、更深入的角度探索研究。



微软亚洲研究院全球研究合伙人韦福如

邱锂力：我们需要不断学习新的知识，尤其是跨领域的知识，因为人工智能正在飞速发展，并广泛应用于医疗、社会科学等不同领域。对于人工智能研究者而言，持续学习和快速学习的意愿和能力将成为至关重要的特质。

现场观众：我的专业是软件工程，我会接触到黑盒测试和白盒测试等概念，而深度神经网络这样的人工智能模型更像一个黑盒。我们一直在关注人工智能的负责任和可靠性问题，那么，我们现在能够在多大程度上信任人工智能模型？是否需要先理解人脑在生物层面的工作机理，才能创造出更完善的人工智能？

韦福如：我们现在还无法真正理解大语言模型的运行原理，但随着研究的深入，这些机制将逐渐为人们所了解。虽然我们不能完全相信模型的输出结果，但我认为它们会越来越可靠。当然，在做出重要决策时，还是需要将人工智能和人类智能相结合，人类将发挥非常重要的决定性作用，而人工智能将成为一个越来越强大的工具。

邱锂力：了解人脑的生物学原理，会有助于我们更深入地探索人工智能。例如，目前的神经网络主要是前馈的结构。但是神经科学家发现，人脑的神经元不仅有前馈，还有后馈、抑制、促进等多种信号传输机制。也就是说，神经元的信号传递并不是多个独立一对一的输入结构，而是各个输入之间存在依赖关系，涉及多种反馈机制的综合作用。

因此，我的同事们设计了融合多种反馈机制的新型网络结构 CircuitNet，相比传统的前馈网络结构，CircuitNet 体量更小，性能更好。他们的研究经验表明，脑启发研究可以帮助我们改进传统的 Transformer 或其他神经网络结构，设计出更优越的下一代神经网络。

国吉康夫：我认为没有百分之百确定的神经网络系统，因为它是一个概率系统。人工智能的可靠性和可解释性是非常重要的问题，而随着技术的发展，这些问题也在得到改善。我们也不应该期望有百分之百的正确性，系统会犯错，就像人也会犯错一样，但有时候错误是可以理解的。目前的人工智能会犯一些很荒谬的错误，这才是真正的问题。



东京大学新一代智能科学研究中心负责人、信息科学系研究部智能机械信息技术学教授国吉康夫



扫描二维码观看 AI Forum 2023 大会视频回放





我今年 25 岁，与微软亚洲研究院同龄

这一年，我们 25 岁！

25 年，一个世纪的 1/4。25 岁这一年，对于有些人而言或许面临着就业；有些人可能正处在科研生涯的关键节点；还有的人可能还在发懵，怎么一晃就 25 岁了……但对于大部分“过来人”来说，25 岁真年轻啊！

面对人生的十字路口，我们都有对未来的期许和对不确定性的迷茫。当 Z 世代的 25 岁迎面撞上 60 后、70 后、80 后 25 岁时的青葱岁月，我们会从时光的汇流中发现什么？在那些你熟悉和崇拜的研究员眼中，25 岁的年轻人是什么样子？他们又会给当年 25 岁的自己什么样的建议？



扫描二维码观看视频

我今年 25 岁，与微软亚洲研究院同龄

25 岁，风华正茂，是无数年轻人出发去改变世界的年纪，而微软亚洲研究院影响未来的脚步从未停止。

没有人永远 25 岁，但永远有人正 25 岁。你还记得你 25 岁时面对的是什么呢？那时的目标是否已经变成了现实？那些资深的研究员们，在他们最初踏入研究院时，对这里的印象又是怎样的？在研究院的日日夜夜里，哪些是属于他们最闪亮的记忆？从过去到未来，我们看见了研究院的哪些成长，又对研究院的未来有着怎样的期许呢？



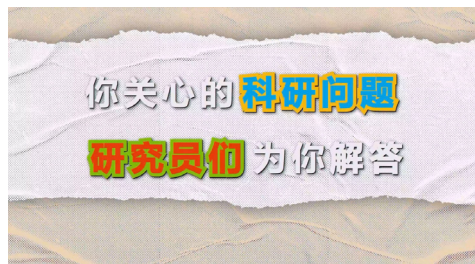
扫描二维码观看视频

如何在资源导向的科研大环境中寻找机会？

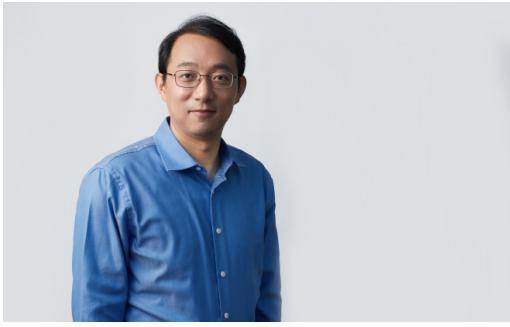
做科学研究从来不是一件简单的事情，在最前沿和最尖端的技术领域内挑战只会更多。

每个踏足科研的同学都有影响世界的理想，但也难免受到现实的挫败：也许是一次实验出错，也许是迟迟不能定下的选题，也有可能同时面对难以达成的论文要求和充满困惑的就业选择。每个微小的困难，在我们自己亲身走过时都显得分外沉重。

在这样的关口，那些轻舟已过万重山的研究员们或许能为你提供一些有益的建议和实用的经验：如何在资源导向的科研大环境中寻找机会、怎样快速更换研究方向、如何保持长期的科研热情……你需要的答案就在这里。



扫描二维码观看视频



周礼栋

微软全球资深副总裁、微软亚洲研究院院长

“二十多年来，微软亚洲研究院始终秉承开放、积极的心态，致力于打造自由、平等、可持续的科研协作环境，让分工、协调、合作链环上的每个人都成为新的发现与贡献的核心主体，为各种创造性想法的星星之火提供燎原之势的催化剂。

一个创新型组织的成长是不断拓展视野并承担更大社会责任的过程。微软亚洲研究院从创立伊始就持续与国内外计算机科研机构展开深度合作，携手进步，共同发展。在面对当下可持续发展、碳中和、医疗健康等人类社会亟待解决的关键问题时，微软亚洲研究院将守正创新，践行所有有利于激发创新力的原则，大胆接受和改造各种新的范式，与各界伙伴共同推动计算技术的跨界融合发展。”

关于微软亚洲研究院

微软亚洲研究院成立于1998年，是微软公司在亚太地区设立的研究机构。通过来自世界各地不同学科和背景的专家学者们的鼎力合作，微软亚洲研究院已经发展成为世界一流的计算机基础及应用研究机构，致力于推动整个计算机科学领域的前沿技术发展，将最新研究成果快速转化到微软的关键产品中，并且着眼于下一代革命性技术的研究，助力公司实现长远发展战略和对未来计算的美好构想。

作为微软研究院全球体系的一员，微软亚洲研究院拥有广阔的国际视野，同时扎根中国，辐射亚洲，通过融合东西方创新文化的精髓，以高度的社会责任感，持续开展有影响力、有温度、面向未来的基础科学研究和技术创新。微软亚洲研究院始终秉持相互信赖、相互尊重以及开放合作的理念，承诺与高校和科研机构开展持久而有效的合作，激发创新潜力、推进行业发展。

微软亚洲研究院倡导对技术进步怀有远大抱负，推崇富于冒险的极客创新精神，鼓励研究人员拓展研究的深度与广度，跨越计算机领域的界限，把视野拓展到解决具有广泛社会意义的问题上：提高人类的知识水平，推动基础研究的发展；增强人类的创造力和成就；培育有韧性、可持续的社会；支持健康的全球社会；确保技术值得信赖，让每个人都可以受益。



扫描二维码观看视频介绍

微软研究院全球布局





微信



知乎



小红书



电话: 86-10-59178888

网址: <http://www.msra.cn/>

微博: <http://t.sina.com.cn/msra>