



2023年 第1期

# 江苏省计算机学会通讯

COMMUNICATIONS OF THE JSCS



- ChatGPT: 语言模型的大爆发发展趋势与挑战
- 审计大数据智能分析关键技术与应用

# 江苏省计算机学会常务理事单位

## 南京工业大学计算机科学与技术学院

计算机科学与技术学院前身源于南京化工学院 1994 年成立的自动化与计算机系，2001 年原南京化工大学与南京建筑工程学院合并组建了南京工业大学，2002 年由原两校的相关专业成立了信息科学与工程学院，2008 年更名为电子与信息工程学院，2015 年因学校学部制改革更名为计算机科学与技术学院。学院现有计算机科学与技术系、电子信息工程系、通信工程系、基础教学部和实验教学中心。

学院拥有一支结构合理、以青年教师为主的高水平师资队伍。现有教职工 115 人，其中专任教师 83 人。教职工中，具有正高级职称 11 人、副高级职称 46 人，具有博士学位的 51 人，拥有国家级人才 1 人，省部级高层次人才十余人。

学院现有计算机科学与技术、电子信息工程、通信工程、人工智能四个本科专业，其中计算机科学与技术专业、电子信息工程专业为国家级一流本科专业建设点，通信工程专业为省级一流本科专业建设点，人工智能专业为 2020 年新申办专业。计算机科学与技术、电子信息工程专业通过了教育部专业认证。目前，各专业本科生 1620 余人，研究生 260 余人。学院始终坚持立德树人根本任务，紧紧围绕学校建设国内一流国际知名创业型大学建设目标，强化教育教学管理，积极探索创新、创业和创造型人才培养模式，努力培养会认知、能做事、善合作、谋发展的复合型高素质人才。近年来，学院学生在全国大学生电子设计大赛、ACM 邀请赛、FPGA 邀请赛等重大赛事中获得数十项全国一、二等奖。

学院现有计算机科学与技术一级学科硕士点（学术型）和“电子信息”类别硕士点（专业型），共建“智能制造技术与工程”交叉学科博士点。学科方向主要为网络与云计算、智能感知与信息处理、数据科学与智能计算、生物信息与医疗大数据、软件理论与方法和人工智能等。建有网络与云计算技术研究所、计算机应用研究所、人工智能研究所、信号处理与智能传感研究所和数字城市与智能建筑研究所等。学院坚持理论研究和实际应用并重，坚持与学校化工、生物、安全等优势学科的交叉研究和协同创新，坚持产学研结合，与华为公司、江苏省未来网络创新研究院等建立了深度的合作关系。近几年，承担了国家自然科学基金、江苏省重点研发计划、江苏省自然科学基金等国家和省部级项目 40 余项，科研到款 8000 多万，授权发明专利 60 余项。一些研究成果发表在本学科的高水平期刊或会议上，获得中国石油和化学工业联合会科技进步奖一等奖等科学技术奖 8 项。





# 江苏省计算机学会通讯

## COMMUNICATIONS OF THE JSCS



### 顾问委员会

主任：周志华  
副主任：武港山 耿新 胡江溢  
李斌 夏士雄 李凡长  
陈兵 詹永照 程光  
委员：罗军舟 肖亮 申富饶  
陶先平 吉根林 胡孔法  
张道强 黄强 邓建明

### 编委会

主编：路通  
副主编：聂长海 张洁  
编委：钱柱中 游辉敏 石克

地址：中国江苏省南京市栖霞区仙林  
大道 163 号

邮编：210023

电话：025-89680909

邮箱：jscs@nju.edu.cn

### 学生记者团专栏

01 | ChatGPT：语言模型的大爆发发展趋势与挑战

### 政策解读

09 | 宿迁市人才科技创新政策

### 学术交流

15 | 云存储中数据完整性审计技术研究 | 周磊

### 会员风采

27 | 在“边缘智能”的科研道路上勤勉耕耘 | 许小龙

32 | 在“云端融合”的科研与实践道路上砥砺前行 | 东方

### 科学普及

35 | 啥?! 能打败 GPT4 的“强人工智能”来了?

### 科创成果

42 | 审计大数据智能分析关键技术与应用

### 学会动态

14 | 江苏省计算机学会第八届常务理事会第八次会议在线成功召开

26 | 江苏省计算机学会获 2022 年度“科创江苏”专项行动计划结题  
认定



# ChatGPT：语言模型的大爆发、发展趋势与挑战

## ——江苏省科技工作者及高校学生的专门采访与问卷调查

### 第一期学生记者团

夏烨（南京农业大学） 王益镛（南京农业大学） 杜雅莉（南京大学）

季钰超（南京大学） 薛志坤（南京大学） 王璇（南京大学） 熊浚丞（南京大学）

关键词：AIGC 自然语言处理 ChatGPT

ChatGPT 是一种基于 Transformer 结构的大型语言模型，已经在多个领域展现出强大的能力，如问答、对话生成、文本生成等。但是，随着模型规模的不断增大和应用场景的不断拓展，模型所需要的计算和数据资源也越来越大，同时还面临着一些实际应用中的挑战和难题。为了更好地探究语言模型的发展方向和未来趋势，本文将从多个角度深入探讨 ChatGPT 等生成式 AI 模型（AI generated content, AIGC）的发展现状、面临的挑战以及未来的发展方向，包括问卷调查、专家采访、问题思辨等。我们希望本文能够为读者提供深入、全面的了解，并对未来研究和应用方向提供一定的参考和启示。

### ChatGPT 和语言模型：公众对未来应用和发展的看法调查

为了深入探究大众对于 ChatGPT 的认知和使用体验，我们对江苏省内高校学生开展了一项问卷调查，并收集了 236 份有效问卷。该调查旨在通过设计问卷形式，收集有关 ChatGPT 易用度、泛用度、可信度等方面的信息，同时还收集了用户对于 ChatGPT 的优点、不足和改进建议等反馈，以便更全面地了解大众对于 ChatGPT 的认知和使用情况。

在易用性方面，我们将着重关注 ChatGPT 的对话体验、交互流程、响应速度、交互设计等方面。易用的交互方式是用户与 ChatGPT 进行沟通和交流的重要途径，因此对于用户来说，交互方式是否清晰易懂、流畅自然、反应迅速是相当关键。

根据问卷结果显示，在关于“Q3-1 您最喜欢 ChatGPT 的什么特性”的问题中，73.9% 的受访者选择了文本生成丰富度高。这表明，大部分用户对于 ChatGPT 在生成文本方面的表现比较满意，认为它可以生成具有一定逻辑性和连贯性的文本内容，并且能够提供相对准确、有用的信息和答案。此外，有 43.5% 的用户认为 ChatGPT 的对话式的使用方式简单易用。这表明，在 ChatGPT 的设计中，对话式的交互方式对于用户来说是比较友好和易用的，用户可以通过简单的指令或问题与 ChatGPT 进行交互，从而方便地获取所需信息。

Q3-1 您最喜欢 ChatGPT 的什么特性? [ 多选 ]

回答速度	0.13
回答准确度	0.13
文本生成丰富度	0.739
语言理解能力	0.391
情感理解能力	0.174
交互友好	0.261
对话式的使用方式	0.435

在泛用性方面，我们将关注 ChatGPT 在不同领域、不同语言和文化背景下的适应能力。我们希望了解用户在使用 ChatGPT 时，是否能够获取到对于自己所处领域和文化环境具有价值的信息，以及 ChatGPT 是否能够准确地理解和解答不同语言和文化背景下的问题。同时，我们也将关注 ChatGPT 的可扩展性和可定制性，以便满足不同用户的需求和使用场景。

根据问卷结果，在关于“您觉得 ChatGPT 可以在工作中为您提供哪些方面的帮助”的问题中，多数用户认为 ChatGPT 在工作中可以提供自然语言生成、搜索与知识获取方面的帮助，分别占 71.3% 和 69.6%。同时，也有不少用户认为 ChatGPT 可以提供多语言翻译和内容创作与编辑方面的帮助，分别占 65.2% 和 63.2%。根据问卷结果，可以看出 ChatGPT 在工作中的应用场景较为广泛，包括自然语言生成、搜索与知识获取、多语言翻译、内容创作与编辑等方面。用户对于自然语言生成和搜索与知识获取的需求较为强烈，而多语言翻译和内容创作与编辑方面的需求也是不可忽视的。ChatGPT 在不同领域和行业的应用前景广泛，具有较高的商业和社会价值。

Q2-7 您觉得 ChatGPT 可以在工作中为您提供哪些方面的帮助? [ 多选 ]

自然语言生成	0.713
多语言翻译	0.652
内容创作与编辑	0.632
搜索与知识获取	0.696

在可信度方面，我们将关注用户对于 ChatGPT 的可靠性、精度和安全性等方面的认知。我们希望了解用户对于 ChatGPT 生成文本的可靠性、准确性和真实性的评价，以及用户对于



ChatGPT 保护用户信息和隐私的安全性认知。同时，我们也将关注用户对于 ChatGPT 的数据来源和算法透明度的评价，以便更好地理解用户对于 ChatGPT 的信任和认知程度。

根据问卷结果，在关于“您认为 ChatGPT 在哪些方面的表现需要优化？”的问题中，用户认为 ChatGPT 在可信度方面仍有提升的空间。有超过一半的用户认为 ChatGPT 生成文本的可靠性和准确性需要优化，占比为 52.2%。此外，有超过三分之一的用户认为 ChatGPT 的数据来源和算法透明度不够，占比为 32.3%。另外，有约 63.5% 的用户认为 ChatGPT 需要优化用户信息和隐私的保护措施。这些结果表明，用户对于 ChatGPT 的可信度和安全性方面有一定的疑虑和不满，认为 ChatGPT 在生成文本的可靠性、数据来源和算法透明度、以及用户信息和隐私保护方面都需要进一步提升。

---

Q3-2 您认为 ChatGPT 在哪些方面的表现需要优化？ [ 多选 ]

---

生成文本的可靠性	0.522
用户信息和隐私的保护措施	0.635
数据来源和算法透明度	0.323

---

根据本次问卷调查结果，用户对于 ChatGPT 的认知程度和使用体验总体表现良好。在易用性方面，用户认为 ChatGPT 的对话体验、交互流程、响应速度、交互设计等方面都较为友好和优秀。在泛用性方面，用户认为 ChatGPT 在自然语言生成、搜索与知识获取、多语言翻译、内容创作与编辑等方面具有较强的应用能力和商业价值。然而，在可信度方面，用户对于 ChatGPT 的生成文本的可靠性、准确性和安全性存在一定的疑虑和不满，也认为 ChatGPT 在数据来源和算法透明度方面还有进一步提升的空间。

### 探讨语言模型发展：解读 ChatGPT 的趋势和挑战

为了深入了解领域研究者和行业从业者对于 ChatGPT 的看法，我们采访了南京大学计算机科学与技术系黄书剑副教授、南京大学计算机科学与技术系许蕾副教授、WSDM Cup 2023 VQA 冠军高圣沂同学。我们从 ChatGPT 爆火之由、ChatGPT 或将面临的挑战、ChatGPT 与软件工程三个方面进行了采访，以全面了解 ChatGPT 在自然语言处理领域的应用、研究和发展方向。

#### ChatGPT 爆火之由

##### Q1：超大体量模型的优势是什么，为什么能够带来更好的语言理解和生成能力？

高圣沂：超大体量模型主要优势来自两方面：大体量的模型参数和训练数据。其中，大体量的模型参数给予了模型更为强大的语言表示能力和上下文理解能力，这也是 ChatGPT 等大模型的表现远超过去的小模型的一大因素。另一方面，如此大规模的模型参数必然需要大量的数据进行训练。这些数据更加多样化、复杂化，涵盖的知识面也更加广泛，这使得模型能够学习到更多的语义信息。

##### Q2：在您看来，ChatGPT 模型的成功因素是什么？开源、应用场景需求、模型体量、技

术创新等哪些方面是重要的？

**高圣沂：** ChatGPT 的成功主要得益于模型体量和应用场景需求。巨大的模型体量配合大量的训练数据使得 ChatGPT 学习到了远超一般模型的能力，其突出的表现也一下子吸引了人们的眼球，促成了其成功。同时，ChatGPT 可以适应很多的下游任务，语言在实际应用场景中又是十分常见的，比如搜索引擎等等，这也促使它不会成为一座空中楼阁，只有投入到实际应用中才能够取得人们不断的关注。相对来说，创新方面我认为并不突出，因为本身其结构依然是基于 Transformer 的，主要在于模型规模的扩大。同时，开源方面目前 ChatGPT 也并未开源，因此这一因素并不好评判。

### ChatGPT 与软件工程

**Q1：ChatGPT 在软件工程领域有哪些应用和发展前景？**

**许蕾：** 就我们所在的研究领域而言，ChatGPT 可以用于自动编写代码、处理代码相关的文档以及智能编程辅助等。首先，ChatGPT 可以用于自动编写代码。它可以通过与程序员进行对话来理解程序员的需求甚至意图，并据此生成代码片段或代码框架。这可以大大提高编程效率和代码质量，特别是在快速原型开发和重复性编程任务方面，比如前端开发中，能根据界面效果图自动生成对应的代码。其次，ChatGPT 可以用于自然语言处理（NLP）任务，如文本生成、机器翻译、情感分析等，在软件工程中有着广泛的应用。开发者可以只专注于业务逻辑，而自动生成代码注释、自动生成开发文档等任务可以由 ChatGPT 完成。再次，ChatGPT 可以用于智能编程辅助。它可以通过与程序员对话来提供编程建议（如 API 推荐、代码补全）、错误修复和重构建议等。这可以帮助程序员更快地解决编程问题，提高编程效率和代码质量。总之，ChatGPT 在软件工程领域有着广泛的应用和发展前景。随着人工智能技术的不断发展，ChatGPT 将成为软件开发中不可或缺的工具之一。

**Q2：在智能编程辅助上使用 ChatGPT 会面临哪些挑战？**

**许蕾：** 作为大语言模型的代表，ChatGPT 在预训练过程中便使用海量代码语料库来提升代码理解生成能力，因此能够为代码自动补全、API 推荐等智能编程辅助方向提供帮助。然而，在智能编程辅助领域，ChatGPT 的使用仍面临一些挑战。首先，ChatGPT 作为文本大模型，其处理主体更多在于文本而非代码，关于代码间的语义信息（如数据流、控制流、依赖链等）缺少相应的分析、理解；其次，多数编程辅助任务往往需要实时更新代码上下文场景，而 ChatGPT 知识更新不及时，可能缺乏现实编程场景适应能力；此外，ChatGPT 的输出缺少稳健性，无法保证生成结果的准确性。因此，智能编程辅助虽然能提供辅助能力，但无法被重点依赖。综上所述，ChatGPT 能够为智能编程辅助提供推荐结果，但仍需用户具备对编程任务的自我思维、逻辑知识和判断能力，从而才能够起到事半功倍的效果。

### ChatGPT 或将面临的挑战

**Q1：您是否认为大型语言模型有能力自发产生“道德”？如果不行，人类应该如何对其进行限制？**

**黄书剑：** 个人认为，大型语言模型目前还没有能力产生道德。因为道德是人类理性思考



的结果，是在社会利益冲突中形成的被广泛认可的行为准则。现阶段的人工智能是单个模型独立学习，没有涉及到利益冲突或是成本收益的考量，因此很难在这种情况下产生我们所理解的道德。它的看似具有感性的回答只是从学习的数据中获取的模式，而没有独立的情感和理性思考能力。因此，产生道德还需要进一步的发展。

当然，如果人类希望让 AI 具备某些道德相关的内容，这是可以实现的。例如 ChatGPT 会拒绝回答一些道德上不当的问题，这是使用人类反馈的强化学习方法实现的。这种方法可以建立道德倾向，避免一些法律政策方面的问题，我认为非常重要。

**Q2: 如果考虑到 ChatGPT 的数据很大程度上来源于互联网，如果它被广泛应用于互联网，它的产出是否会反过来作为输入，从而对模型产生负面影响呢？**

**黄书剑:** 确实存在这样的问题。我个人是做机器翻译的，早在很久以前，机器翻译领域就已经存在了这样的问题。比如，我们在互联网上开放了一个翻译引擎 A，虽然我们不主动爬取引擎 A 的结果，但其他网站可能会使用它的翻译结果，最终我们仍会获取到其产出，这就会存在问题。因为翻译的结果可能并非完全准确，导致产生一些负面影响。虽然一般我们可以收集大规模的数据，但对数据的清理仍然非常重要。

**Q3: GPT4 已经支持了一定程度上的多模态输入，但其计算资源消耗也在增加。您认为大型语言模型未来会朝着专业化的方向发展，还是会继续提高其通用性？也就是说，它们将成为高级专家系统或通用系统的角色？**

**黄书剑:** 这是一个非常好的问题。不同的研究人员可能会站在不同的角度看待。从通用人工智能的角度来看，我认为 OpenAI 仍将继续开发更大型的模型。即使 GPT 已经很大了，他们仍会追求突破极限。从智能的角度来看，大而强是不可避免的趋势，但同时大型模型也会带来各种问题。其中一个垄断问题。例如，OpenAI 在开发 ChatGPT 时投入了难以想象的资源。至少在当前阶段，大型模型很有可能是寡头，而这对应用相关发展不利。另一个问题是个性化。尽管模型已经非常智能，但在某些任务上仍不够好。我们很难直接对大型模型进行个性化调参。因此，为了满足特定的服务，可能需要采用相对小而精的模型。

在以上采访中，我们了解了 ChatGPT 在自然语言处理领域的应用、研究和发展方向，以及在软件工程领域的应用和发展前景。采访中的专家们认为，ChatGPT 的成功主要得益于其大体量的模型参数和训练数据，这使得它具备了更强大的语言表示能力和上下文理解能力。此外，ChatGPT 还可以用于自动编写代码、处理代码相关的文档、智能编程辅助等多个软件工程领域任务，并具有广泛的应用和发展前景。

然而，ChatGPT 的使用仍然面临一些挑战。在智能编程辅助领域，ChatGPT 的处理主体更多在于文本而非代码，其输出缺乏稳健性，无法保证生成结果的准确性，因此无法被重点依赖。另外，大型语言模型可能会带来一些问题，如垄断和个性化问题。

最后，采访中的专家们认为，虽然大型语言模型有可能产生一些“道德”，但目前还不具备这样的能力，而限制大型语言模型的应用需要综合考虑人类利益和技术发展等多个方面。在未来，大型语言模型可能会朝着专业化的方向发展，同时也需要平衡其与人类社会的关系。



## 深度学习模型的未来发展：探讨“小而精”和“大而强”的路线选择

深度学习模型的发展方向是一个备受争议的话题。在过去的几年里，随着深度学习技术的快速发展和大数据时代的到来，越来越多的研究人员倾向于使用大规模的神经网络模型来解决各种复杂的任务，如自然语言处理、计算机视觉、语音识别等。这些模型通常由数百万或数十亿个参数组成，需要大量的计算资源和数据支持，称为“大而强”。

然而，随着深度学习技术的逐渐成熟和普及，一些研究人员开始注意到“小而精”模型的潜力。这些模型通常由数百个或数千个参数组成，可以在低功耗设备上运行，并且能够在一些特定任务上达到与大型模型相当的性能水平。同时，这些模型具有更快的训练和推理速度，更易于部署和维护，逐渐成为研究人员的新宠。

在这一部分，我们将着重探讨深度学习模型的发展方向，从“小而精”还是“大而强”两个角度展开思辨。

### 正方观点：“大而强”是深度学习模型发展的必然趋势。

深度学习超大体量模型体现在大体量的模型参数和训练数据中，这也使得这些模型具有更丰富的语言知识储备、更精细的特征提取能力、更好的泛化能力以及更强的可迁移性，从而能够更好地应对各种语言任务和应用场景。正因如此，深度学习大模型的研究和应用已经得到了广泛的关注和发展。例如，在自然语言处理领域，BERT、GPT-2 等大型预训练模型已成为了主流，它们可以在各种下游任务中取得领先的性能。

ChatGPT 是一种基于大规模语言预训练的自然语言处理模型，它可以回答多个领域的问题，甚至能完成写作、编程等专业操作，因此，我们有理由相信，随着模型规模的大幅度提高，模型的精度会进一步提升，获得更突出的效果。事实上，ChatGPT 有着远比它现在看起来更丰富且广阔的应用前景，它的成功让人们看到了超大模型实现通用人工智能的潜力。在现有的特定领域的人工智能模型研究上，我们已经取得了不错的成果，人工智能也已经具有了较强的能力，但在特定领域拓展到解决其他问题上，就有较大的难度，也就是缺少将不同算法统一起来的大模型。用一个比较大的模型来完成多种功能，这是人工智能的一个主流的发展趋势。这就与通用人工智能发展有关了，通用人工智能（AGI）是具有一般人类智慧，可以执行人类能够执行的任何智力任务的机器智能。通用人工智能是一些人工智能研究的主要目标，也是科幻作品和未来研究中的共同话题。在电影《流浪地球 2》中与刘培强对话时期的 MOSS 就是相较于 ChatGPT 更泛化的人工智能。

想要像 ChatGPT 这样的模型一样取得这么好的效果，我们就需要发展更大规模的模型，这也是必然的趋势，大模型的表现已经使得传统的小模型黯然失色。因此除了一些特定任务，在那些基本任务上必然会不断更新模型，以更大的体量实现更强大的能力，能够适应环境且在知识和资源不足的情况下继续工作。在其发展过程中，因为并没有任何不可能的条件，所以我们有理由相信，也有理由展望更大更好的模型的出现，带来更多技术上的突破。

### 反方观点：“小而精”是深度学习模型发展的正确方向。



深度学习模型的发展方向并没有绝对的对错之分，但在实际科学研究中总有一个主流的方向。ChatGPT 的成功代表大型通用模型将成为未来研究的热点，但我相信“小而精”的专业化模型仍将会是重要的发展方向。

我们先要对“小而精”的专业化模型进行解释说明：其中“小而精”是一种深度学习模型设计策略，其主要思想是在保持高精度的前提下，通过减少模型参数和计算量来实现模型的轻量化。而专业化则是指针对特定任务或场景进行模型的设计和优化，以达到更好的性能表现。当然 ChatGPT 的成功证明了大型通用模型也能很好的解决特定场景的专业化问题，但基于以下几个原因，我们相信轻量化、专业化的模型仍将保持相当的生命力。

首先轻量化模型具有在模型训练和部署上的优势：通过减少模型的参数和计算量，从而减少计算成本，对于在低功耗设备上部署模型的场景来说，这一点尤为重要。而且计算成本的降低为模型的推广降低了门槛。通常情况下，能被广泛应用并能形成一定研究生态的模型往往有着更大的发展空间，即在一定程度上，相对较低的部署条件为模型发展提供了助力。除此之外，轻量化模型还能降低对大规模语料库数据的依赖，避免训练数据数量成为模型瓶颈，同时数据处理难度的降低也为提高模型的专业性与可信度提供了可行路径。

其次，在理论研究方面，大型模型的高复杂度与大量参数虽然带来了优秀的性能，但也给模型的可解释性研究带来了巨大的困难。理论研究是深度学习发展的关键领域之一，对模型原理理解的滞后将对模型的决策信任度产生负面影响，这通常意味着模型落地应用成本的增加。因此，轻量化模型的发展对于研究人员基于较小模型探索通用性理论具有不可或缺的价值。此外，轻量化模型的发展对于探索更高效的模型结构与更高效的训练算法也有着重要意义。

最后，从实际应用的角度来看，专业化模型在当前的应用场景中仍然占据着重要地位，专业化模型可以提高模型性能、减少计算资源消耗、提高模型可解释性、推动深度学习在特定领域应用的发展。此外，深度学习模型的社会价值大小往往取决于其是否能真正落地，完成与实际产业的对接，而专业化正是模型实际落地的必然环节。当然，目前主流的专业化方案大多是针对特定实际需求场景来设计专业化模型，但伴随着大型通用模型的发展，模型专业化的形式可能会转变为对通用模型进行修饰和特化。总体来说，模型专业化仍将是模型实际应用不可或缺的重要环节。

综合而言，未来深度学习模型的发展方向应该是“小而精”和“大而强”并存。在大规模数据和高性能计算资源的支持下，大型神经网络模型仍然具有广泛的应用前景，尤其是在复杂任务和领域上。同时，随着边缘计算和物联网的发展，小型神经网络模型的应用也将逐渐扩展，成为解决低功耗设备和实时应用的重要手段。因此，我们需要在“小而精”和“大而强”之间找到平衡，不断创新和探索新的技术和模型，以满足不同场景和需求的应用需求。

## ChatGPT 和 AIGC 的未来构想：人工智能与人类的共生之路

21 世纪以来，伴随着信息科学技术的发展，在移动互联网、大数据、神经网络和深度学

习等新技术的推动下，人类正加速迈进崭新的人工智能时代。随着脑科学、认知学习理论及大量交叉性复杂性学科的迅猛发展，大数据技术、深度学习算法以及计算机算力的突破，再加上政府、企业等的外在推动，人工智能已逐渐渗透进人类社会的各个领域，并助力应用领域的全新变革。《全球人工智能发展白皮书》将人工智能定义为使用机器代替人类实现认知、识别、分析、决策等功能，模拟人的某些思维过程和智能行为而创建的模型。为 ChatGPT 这类工具提供动力的正是蓬勃发展着的人工智能，而人工智能的未来发展也将会进一步加快。

在未来，人工智能或许会有着更为广泛的应用空间和更为个性化的智能服务。人工智能将被广泛应用于医疗、教育、交通、制造业等各个领域，推动 AI 问诊、智能教学、无人驾驶汽车、智能家居、智能城市等的建设和发展，为人们提供更多的便利和服务，促进社会进步和人类生活的改善。人工智能还将更好的对人们个性化的需求和偏好进行识别，从而能够为每个人提供更符合其切实需求的服务，与人们一起协同办公，成为人们日常工作和生活中的得力助手。

人工智能技术的不断发展和应用必将改变我们的生活方式和社会结构。为了适应以人工智能为驱动的社会，我们需要不断学习新的技能，从而增强自身核心竞争力。人工智能技术势必会在某些领域取代人类的工作，但同时也会催生出新的工作需求，所以我们应当重视自己的自我价值和独特贡献，增强自信心和幸福感，同时不断提高创造性思维、批判性思维、协作能力和情绪感知力等机器难以具备的思维和能力。

目前在和 ChatGPT 交互的过程中，我们可以发现人工智能模型并不擅长理解人类请求的上下文，在回复人类的请求时，可能会给出错误的答案或编造的信息。也许在未来，这个问题会在技术人员的研究下得到有效的修复，但仍然需要我们在使用人工智能时具备基本的辨别回答可靠性的能力。除了对可靠性的担忧，人工智能的算力和学习速度远超人脑，如果其被恶意利用，将为人类社会带来巨大的威胁。因此我们在利用人工智能时，应当更加注重伦理和安全，避免出现不良后果和风险。我们需要大力推动政府、企业、研究机构、社会组织等多方合作和信息共享，共同推动人工智能技术的可持续发展；制定相关法律法规和伦理规范，加强对人工智能的监管和管理，确保人工智能的合法、公正、安全和可信；制定人工智能应用的安全标准和规范，确保人工智能应用的开发、测试、部署、运行等环节都合规；强化数据安全和隐私保护，避免数据被不当使用和滥用，保护个人隐私和商业机密；加强人工智能的透明度和可解释性，让人们能够理解和控制人工智能的行为和决策，减少因人工智能决策带来的风险和误解。

总之，人工智能技术的发展需要全社会的共同努力和参与，只有积极探索和应用人工智能技术，并加强伦理意识和监管，才能更好地利用人工智能，为人类创造更多的价值和福祉。



# 宿迁市人才科技创新政策

## 1、宿迁科技创新成果

### ①企业创新主体地位不断提升

全市科技型企业竣工验收 238 家，全市高新技术企业达到 534 家，国家科技型中小企业评价入库 1684 家，全市高新技术产业产值占规模以上工业总产值比重达 33%。

### ②科技平台支撑力量持续增强

新认定 3 家国家级孵化器，1 家国家级众创空间，全市省级以上科创载体达到 30 家，新建省级工程技术研究中心 18 家。

### ③科技成果转化再创新高

全市完成技术合同成交额 39.1 亿元，实施高科技成果转化项目 347 项，3 个项目获省重大成果转化专项资金项目支持，获省拨资金 3400 万元，立项数苏北第一，江苏米格新材料公司荣获第十届中国创新创业大赛“创新创业 50 强”“全国总决赛二等奖”，实现历史性突破。

### ④科技创新合作成效显著

全市共签约产学研合作项目 1223 项，获批省“科技副总”227 人，荣获 2021 年“江苏友谊奖”1 项，实现外国人来华工作、居留许可“一站式”窗口联办。

### ⑤科技体制改革进一步深化

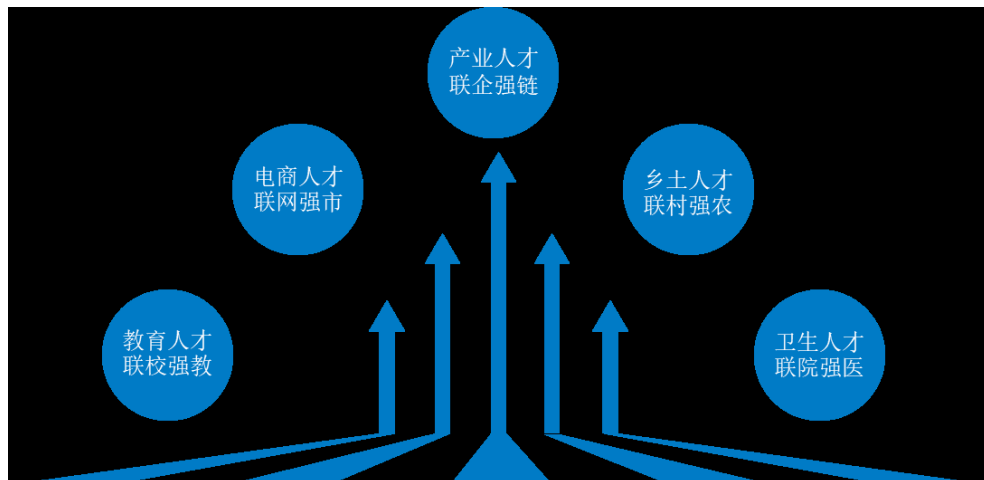
完成“揭榜挂帅”“博士科研启动计划”等 7 条改革事项，率先开展农业科技社会化服务体系省级试点，获准开展科研诚信体系建设省级试点，建成全省首家“产业大脑”创新服务平台，获全市“改革创新奖”三等奖。

## 2、宿迁人才科技政策

宿迁以全市产业发展和民生短板为着力点，推进人才引领服务发展“五联五强”行动，实施更加积极、更加开放、更加有效的人才政策，广泛集聚各方面优秀人才，为经济社会发

展提供有力支撑。

“五联五强”行动方案围绕“十四五”期间发展需要，在人才引进、培养方面双向发力，不断激发和释放人才活力，培养集聚数量充足、结构合理、素质优良的产业人才、电商人才、乡土人才、教育人才、卫生人才队伍，广泛集聚各方面优秀人才，研究出台了系列政策。



### ①千名领军人才集聚计划

支持科技企业和教育、卫生等社会事业领域引进高层次领军人才及团队。经评审认定，给予双创人才 30 万元 -100 万元资助，给予双创团队 100 万元 -300 万元资助，对顶尖人才及团队采取“一事一议”，给予最高 1000 万元资助。

### ②千名拔尖人才培养工程

从全市各类组织中择优选拔一批从事自然科学、社会科学研究、技术开发、推广应用及现代企业管理的本土人才，进行分层培养、分级管理。培养期内，对培养对象发放一定补助，每人每月最高发放补助 3000 元。

### ③人才住房政策

对于来宿就业创业的人才来说，住房是头等大事，宿迁针对住房提供以下三种优惠政策：一是人才“购房券”。按照“来了就发券、买房就兑现”的原则，对新引进的各类高层次人才和紧缺急需人才，全职在宿迁企事业单位工作的，发放 5 万元 -200 万元“购房券”。具体标准为

“两院”院士		>>>>>>	200万元
国家重大人才工程		>>>>>>	100万元
省部级突出贡献专家		>>>>>>	50万元
特级教师和正高职称人才		>>>>>>	30万元
副高职称人才		>>>>>>	15万元



二是生活补贴。新引进到我市产业链企业全职工作或开展自主创业的各类青年人才，根据就业或创业的人才层次给予3万元-50万元不等的人才生活补贴，补贴分三年发放。具体标准为：

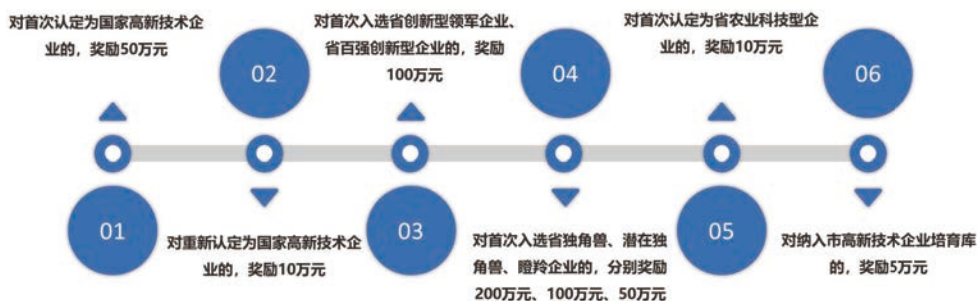


三是人才公寓。我们实施人才安居工程，目前全市建成各类人才公寓30万平方米以上。来宿创新创业的各类人才可就近申请，拎包即可入住。

### 3、科技创新政策

宿迁深入实施创新驱动发展战略，贯彻落实中央和省、市重大决策部署，把科技创新作为引领高质量发展的第一动能，深化“四化”同步集成改革，推动国家创新型城市高水平建设，出台了新一轮科技创新政策，持续培育创新型企业。

#### ①持续培育创新型企业



### ②激励企业开展研发活动

对年度研发费用超过 200 万元但不超过 1000 万元（含）的，按可加计扣除研发费用额的 1%~2% 给予补助；

超过 1000 万元的部分，按可加计扣除研发费用额的 0.5%~1.5% 给予补助；每个企业每年最高补助 100 万元。

对研发费用较上年度增长 10% 以上的，按照增量的 8% 给予补助，每个企业每年最高补助 50 万元；企业前一年度无研发费用投入的，不列入增量补助范围。

对通过评价的国家科技型中小企业，执行企业研发费用加计扣除比例 100% 的优惠政策。对由企业牵头承担的国家科技重大专项、国家重点研发计划、科技创新 2030—重大项目等重大科技项目，按照该项目国家投入的 1:0.5 给予配套支持，单个项目资助额度最高 500 万元。

### ③推动企业建设高水平研发机构

对首次认定为省级企业工程技术研究中心的，奖励 30 万元；

对新认定为省级院士工作站的，奖励 100 万元；

对获得省绩效评估“优秀”等次的工程技术研究中心、院士工作站，按照不超过上述认定奖励金额的 50% 给予依托单位奖励。

对建设国家级博士后工作站、省级博士后创新实践基地的，奖励 20 万元。

对建立离岸研发机构并纳入市备案管理的，最高奖励 30 万元。

对通过省研发管理体系贯标的，奖励 5 万元。

对与省产业技术研究院共建联合创新中心的，按省产业技术研究院支持经费 1:1 给予配套支持。

### ④打造产业重大研发平台

符合宿迁产业发展方向的重大科技创新平台，在建设期内依据创新水平、投资规模等分期分档可给予最高 1500 万元支持；被江苏省产业技术研究院吸纳为加盟专业研究所的，在享受以上政策的基础上再给予最高 500 万元支持。对知名科学家、海外高层次人才创业团队、著名科研机构 and 高校领办的特别重大的专业性、公益性的科技创新平台，以“一事一议”方式给予支持。

对纳入市级备案管理的产业技术研究院开展年度绩效评价，绩效评价“合格”以上的，按其上一年度为宿迁企业开展合同科研、技术转移、科技公共服务等横向服务收入实际到账金额的 15% 给予奖励，最高 100 万元；对评价为“优秀”的，再给予 40 万元奖励。对建成并已实质性运作的新型研发机构，经认定后开展年度绩效考核，参照产业技术研究院予以奖励。对新型研发机构、研发服务型企业、龙头骨干企业、高校院所等有效整合优质创新资源，建设国家级、省级重点实验室或技术创新中心的，分别资助 500 万元、200 万元。

### ⑤支持科创孵化载体建设

对新认定的国家级、省级众创空间或星创天地，分别奖励 30 万元、10 万元。

对列入省级科技创业孵化链条试点的园区，奖励 100 万元；对新认定为国家火炬特色产



业基地的，奖励 100 万元。

对获认定为省级大学科技园的，给予最高 1000 万元资金支持。

对新认定的国家级、省级科技企业孵化器或加速器，分别奖励 60 万元、30 万元；对年度绩效评价为“优秀”的，给予 20 万元奖励。

#### ⑥激励重大科技成果产出

对企业重大创新成果产业化项目给予资助，单个项目最高资助 300 万元。

在市级科技计划中增设“揭榜挂帅”专项、“碳达峰碳中和”专项，保持与省科技计划体系衔接、配套，形成引导全社会开展研发活动的计划体系，提升成果产出效率和能级。

支持与江苏省产业技术研究院共设宿迁市企业科技攻关引导资金，对单个科技攻关项目给予最高 100 万元资助。

对获得国家和省科学技术奖的第一完成单位，分别按 1:4 和 1:2 给予最高 300 万元配套奖励。

#### ⑦激发技术成果转移转化活力

企业引进省内外先进技术成果在本市内转移转化的，按技术合同实际成交额的 5% 给予奖补；企业受让高校院所科技成果单项超过 100 万元且科技成果在本市转化的，按实际支付额的 15% 给予补助，每户每年最高 100 万元。各类技术转移机构为我市企事业单位引进转移转化成果的，按技术合同实际成交额的 2% 给予奖补，单个合同奖励最高 5 万元。技术经纪人（经理人）服务我市企事业单位开展的技术转移活动，按技术合同实际成交额的 1% 给予奖补，单个合同奖励最高 2 万元。高层次人才（团队）携带具有自主知识产权的高科技成果到宿迁创办、领办企业的，鼓励各类投资基金给予优先支持。

#### ⑧构建产业科技服务链

对认定为省技术先进型服务企业的，按照高新技术企业首次认定和重新认定政策给予奖励。

对创成省级产业技术创新战略联盟、省创新联合体的牵头单位，奖励 20 万元；推动科研机构、高等院校的大型科研仪器和实验设施向社会开放，根据服务绩效给予提供单位不高于年度服务收入 20% 的补助，最高 30 万元。对经省评价优秀的，奖励 10 万元。

获评省级科技服务业集聚区（示范区）或特色基地的，给予建设主体 50 万元奖励。

#### ⑨支持高层次科技人才创新

优化基础研究计划（自然科学基金）项目组成结构，面向高等院校、科研院所、医疗机构和新型研发机构，实施博士科研启动项目，每个项目支持市拨经费最高 5 万元。

参加市科技创新创业大赛、市高层次人才创新创业大赛获奖的项目，纳入市级人才、科技计划项目支持范围。

支持我市企事业单位面向经济社会发展重大需求，建设省级外国专家工作室，对新获批省级外国专家工作室的，奖励 10 万元；

支持外国人才申报国家、省友谊奖，对获得国家、省友谊奖的单位，分别奖励 30 万元、10 万元。深入实施科技特派员制度，对科技特派员开展工作给予补助。



### ⑩建立健全多元化科技投入体系

加大财政科技投入，健全财政科技投入稳定增长机制，2022-2025 年市级科技创新专项资金保持稳定增幅，重点用于组织实施科技计划项目、各类科技创新型企业培育和科技平台载体建设奖补以及组织科技创新活动等，加大对基础研究和应用基础研究持续稳定的财政支持力度。加强财政资金和金融手段的协调配合，充分发挥财政资金的杠杆作用，引导金融资金和民间资本进入科技创新领域。设立科技创业投资基金，推动一批引领性强、带动性大、成长性好、具有核心竞争力的高新技术产业和未来产业重大项目在宿落地成长。县（区）要调整和优化财政支出结构，增加财政对科技的投入，提升资金使用效益。



## 学会动态

### 江苏省计算机学会第八届常务理事会第八次会议 在线成功召开

2023 年 1 月 14 日下午在线召开了江苏省计算机学会第八届常务理事会第八次会议。出席会议人员有常务理事会 36 人出席，以及正副监事长、工委负责人、学会秘书处成员列席参加。

会议由金莹秘书长主持。会议首先由周志华理事长致辞，致辞中周志华教授感谢常务理事一年来为江苏省计算机学会做的大量工作，感谢大家长期以来对学会工作的支持，对新年提出美好的愿景！

会议一共 11 项议程。首先审议的是 2022 年度工作报告和 2023 年度工作计划、2022 年度学会财务报告和 2023 年度财务预算，金莹秘书长向常务理事会汇报了 2022 年学会开展的主要工作，从学会的组织机构、党建工作、会员发展、信息化建设、品牌活动、学术会议、科技服务、科学普及、各专委会工作、科技鉴定、科技咨询、竞赛组织以及获得的奖项和荣誉等多方面进行了总结汇报。



# 云存储中数据完整性审计技术研究

——2022 年江苏省计算机学会优秀博士论文奖

作者：周磊，付安民

单位：南京理工大学计算机科学与工程学院

## 论文摘要

数据完整性威胁是云存储数据面临的最为关键性的挑战之一，数据完整性审计技术被认为是检查云存储数据是否被正确存储的有效手段。但目前数据完整性审计技术依然存在一系列核心问题有待解决，比如繁重的标签计算任务不适用于性能受限的本地端、签名密钥泄露威胁审计安全性、不诚实的审计者可能篡改审计结果、数据动态开销过大影响更新效率等。本文围绕云存储中数据完整性审计存在的关键问题展开研究，重点实现云存储环境下单副本审计中公平审计、数据隐私保护、数据动态、密钥暴露抵制，以及多副本审计中授权审计、公平审计、数据隐私保护、数据动态、和可变副本存储等研究目标，旨在提供安全高效的单副本与多副本安全审计服务。数据完整性审计技术方面的相关研究工作，已发表在 TDSC、INS、JNCA、计算机研究与发展、ICC 等国内外一流期刊和会议上。

## 专家推荐语

在云计算和大数据时代的当前背景下，基于云存储的产业与应用需要以云存储数据的完整性不被破坏为前提。因此，确保云存储数据的完整性至关重要。然而，受软硬件故障、人为误操作等因素的影响，云服务器有时会不可避免地篡改或删除存储数据，甚至为了维护自身声誉而隐瞒数据损坏事实。同时，用户可能使用台式机、笔记本、平板电脑、智能手机等多样化的本地端设备接入云存储系统，其中有些设备仅持有有限的计算能力与带宽资源。此外，存储数据的动态更新效率、签名密钥的暴露、审计者的不诚实性等新型需求，也给云存

储中数据完整性审计技术研究，带来了一系列新的挑战和问题。作者从数据拥有者的角度出发，考虑了数据拥有者将数据存储于云数据中心时，面临数据完整性审计的一系列困难，如繁重的标签计算任务不适用于性能受限的用户本地端、签名密钥泄露威胁审计的安全性、不诚实的审计者可能篡改审计结果、数据动态开销过大影响更新效率等。针对上述问题，论文通过使用双线性对、无证书签名、雾计算、区块链等技术，构建了4个满足数据完整性、数据隐私保护、公开审计、数据动态、密钥暴露抵制、可变副本存储、审计公平性等不同安全需求的单副本与多副本审计方案。相比同类技术，论文实现的云存储中数据完整性审计方案，耗费了更少的资源开销，提高了数据完整性审计的在算力贫乏的普通终端设备上的实用性，增强了审计过程抵御恶意攻击、合谋攻击的能力。论文研究工作在学术论文发表方面，取得了显著成效。

### 论文看点

1、针对单副本审计方案难以适用于性能受限本地端的缺陷，提出一个支持标签计算外包与公平审计的单副本审计方案（FCAB）。FCAB通过引入雾节点代替用户完成标签计算与审计交互任务，并结合无证书签名与双线性对技术设计可验证同态标签来实现单副本完整性审计，极大地减轻了本地端的开销。

2、针对单副本审计中存在的签名密钥泄露问题，提出一个支持密钥更新与审计者更换的单副本审计方案（PATCAR）。PATCAR设计了三种分别适用于签名密钥意外泄露、用户身份过期、以及审计者更换场景下的密钥与标签更新机制，并且由云端承担计算复杂的标签更新操作，引入了少量的开销。

3、针对多副本审计中动态数据存储需求，提出一个支持数据动态与隐私保护的多副本审计方案（PDP-D）。PDP-D通过设计一个支持多副本更新的动态结构，实现了数据块内容与序号的双重验证，达到了多副本更新所引入的通信开销与副本数无关的目的。

4、针对多副本审计中所有数据块存储相同副本数导致存储开销过大的问题，提出一个支持可变副本存储与公平审计的多副本审计方案（MDSS）。MDSS通过设计可变副本存储策略与基于无证书的可验证标签，在优化用户存储需求的同时避免了复杂的证书管理开销。面对动态开销与副本数呈线性增长现状，MDSS设计了一种高效的动态结构，使得动态更新所导致的通信开销与副本数无关，极大地提升了多副本更新的效率。

#### 支持标签计算外包与公平审计的单副本审计方案 FCAB

现有云存储完整性审计方案要求用户在本地端为原始文件的所有数据块生成标签，而标签计算往往涉及耗时的群上幂指数与点乘运算，因此难以适用于使用资源受限的本地设备接入云端的用户。此外，现有审计方案总是基于一个诚实的审计者去检测云存储数据的完整性，然而审计者可能仅执行少量计算或与云服务器进行勾结，从而发送错误的审计结果给用户。为此，提出一个支持标签计算外包与公平审计的单副本审计方案 FCAB。FCAB通过引入雾节点来承担标签计算与审计交互工作，有效地减少了本地计算和存储负担。考虑到充当审计者的雾节点在审计过程中可能表现不诚实，用户可以从区块链中提取审计日志来判断历史审计



的真实性，进而实现审计公平性，即确保正确存储数据的云服务器与诚实执行审计任务的雾节点获得服务报酬。FCAB 被证明满足不可伪造性、数据隐私保护和公平审计，在标签计算、数据更新、完整性审计阶段仅引入了较少的计算和通信开销，并且本地端的开销是一个与文件划分块数无关的较小固定值。

支持标签计算外包：在典型的云存储审计方案中，频繁耗时的审计交互过程转交给第三方审计者，用户仅需要在本地为所有数据块计算标签，然而标签计算往往包含耗时的群上指数与点乘操作，因此为所有块计算标签对于本地端设备来说仍然是一个沉重的负担。随着文件大小的增加，计算量将变得更加突出。此外，用户可能使用多样化的本地设备接入云存储服务，其中有些设备仅持有有限的计算能力。因此，现有审计方案难以适用于使用资源受限的本地端设备接入云存储的用户。为降低本地端的开销，在不牺牲审计方案其他性能的情况下，将标签计算安全且有效地委托给其他实体，是降低本地端开销的一种合理方法。作为终端设备与云服务器之间的桥梁，雾节点的计算能力优于终端设备，能有效弥补本地计算能力不足的缺陷。因此，雾节点被引入到本文的审计模型中充当审计者与云服务器进行交互，并协助用户完成标签计算操作。本文利用无证书签名技术，通过在审计模型中引入雾节点，实现了一个面向资源受限本地端的数据完整性审计框架。雾节点帮助用户完成标签计算和审计交互工作。一方面，大部分本地计算都被转移到计算能力更强的雾节点上，本地设备仅需在数据处理和审计阶段执行固定且少量的计算就能验证数据的完整性；另一方面，当本地设备离线时，审计工作仍然可以有效地进行。

提供公平审计服务：作为被引入的外部实体，雾节点很可能在审计过程中仅执行少量计算或者与云服务器进行勾结，从而向用户发送错误的审计结果。因此，用户必须审核雾节点历史审计的真实性，保证只有完成审计任务的雾节点与正确存储数据的云服务器才能获得服务费用，即实现公平审计。本文将量大且需要更新的数据块和标签存储到功能强大的云服务器上，将不能被更改的审计日志存储到具有不可篡改性的区块链上，同时实现完整性审计与公平审计。针对雾节点的每次审计，本文存储一个相应的审计日志到区块链上，其内容是构成该次审计的挑战参数。通过审核审计日志，用户可以判断在历史时间段内雾节点的审计行为是否诚实。为进一步降低日志审核的开销，用户采用抽样检查策略来代替审核所有的审计日志，实现了在某个时间段内无论执行了多少次完整性审计，用户本地端只需消耗等同于一次审计验证的开销，就可以判断历史审计的真实性。实现公平审计的核思想被归纳为：FCAB 规定只有表现诚实的雾节点和云服务器才能收取相应服务费，即只有当云服务器通过了雾节点的验证且雾节点通过了本地设备的检查后，本地设备才认为雾节点和云服务器都是诚实的。当 AuditIntegrity 算法输出 1 时，本地设备认为云服务器通过了雾节点的验证。这时存在两种可能性：1) 雾节点和云服务器都是诚实的；2) 雾节点和云服务器都不诚实。当 AuditIntegrity 输出为 0 时，雾节点将立即通知本地设备，因为它不想承担数据损坏的责任，本地设备将根据当前挑战验证云服务器返回的审计证明。如果证明无效，雾节点被认为是诚实的，云服务器被认为是不诚实的；否则雾节点被认为是不诚实的，云服务器被认为是诚实的。只有 AuditIntegrity 在一段时间内

连续输出 1 后，本地设备才会启动 CheckLog。如果 CheckLog 输出 1，则雾节点通过了本地设备的检查，因此本地设备认为雾节点和云服务器都是诚实的；否则，雾节点和云服务器都被认为是不诚实的。

为降低用户本地端开销并实现公平审计服务，本文结合无证书签名、双线性技术和区块链技术，提出了一种支持标签计算外包与公平审计的完整性审计方案 FCAB。如图 1 所示，在提出 FCAB 的审计模型中，雾节点充当审计者承担与云服务器进行的审计交互工作。此外，用户将处理后的原始文件发送给雾节点，并基于动态结构 P-MHT 与文件签名的一致性保证雾节点基于用户所传的数据生成对应标签，然后雾节点将数据块集合与对应标签集合一起上传给云服务器进行最终存储。在审计挑战中，雾节点验证云服务器上数据的完整性，并将每次审计日志记入区块链。当用户想要验证雾节点是否诚实地完成了历史审计时，用户依靠记录的日志来判断历史审计结果的真实性。与现有审计方案不同的是，不再由审计者产生随机挑

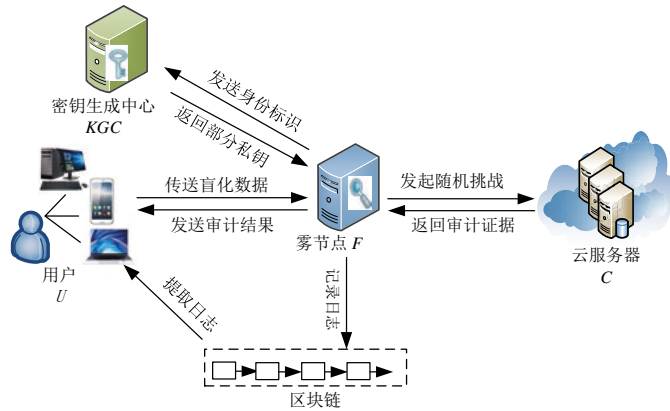


图 1: 方案 FCAB 的系统模型

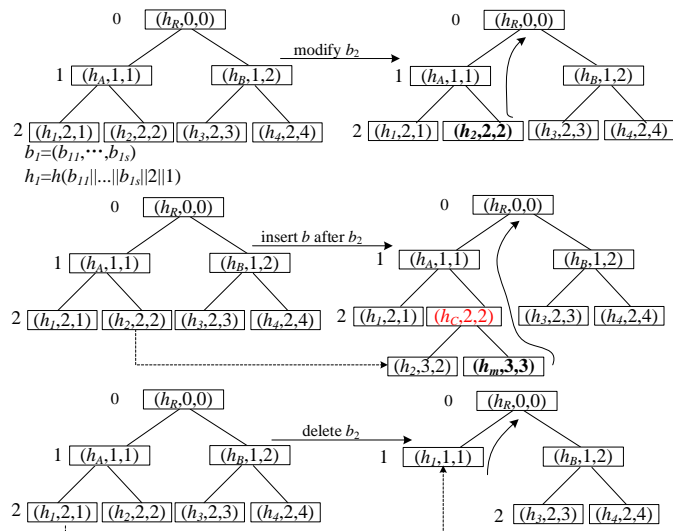


图 2: 动态结构 P-MHT



战并发送给云服务器，利用区块链中不可更改的公共信息来构建完整性挑战，从而避免了雾节点和云服务器伪造挑战的可能性。此外，通过改进传统的 MHT 结构构建新的动态结构 P-MHT，FCAB 可以扩展到动态场景下，进而确保本地设备可以对云存储数据进行动态更新。最后，FCAB 被证实是可证明安全的，且实现了数据隐私保护和公平审计。实验结果证实了 FCAB 的高效性，特别是，在标签生成、完整性审计、以及数据更新阶段，用户本地端仅需要消耗较小且恒定的计算成本与少量的通信开销。

### 支持密钥更新与审计者更换的单副本审计方案 PATCAR

相比于现有完整性审计方案，FCAB 显著降低了用户在本地端的开销，但它与现有审计方案皆面临一个关键的安全问题：签名密钥一旦泄漏，依赖于密钥安全性的审计方案将遭受破坏。此外，现有审计方案均默认有且仅有一个审计者来代替用户进行完整性挑战，然而审计者可能由于被攻陷、被贿赂或资源不足等原因，不能继续为用户提供审计代理服务。因此，本文提出一个支持密钥更新与审计者更换的单副本审计方案 PATCAR，它能有效地抵制签名密钥暴露。PATCAR 通过引入雾节点来充当用户的审计者，以弥补本地计算力不足的缺陷。考虑到签名密钥由于暴露导致基于该密钥的所有签名无效的问题，PATCAR 允许云服务器将这些标签有效地转化为基于新的签名密钥的标签，同时在整个审计过程中云服务器无法获取新的签名密钥。此外，当在职雾节点出于主动或被动原因不能继续提供审计服务时，PATCAR 进一步实现了审计者可更换需求，保证了一个新的雾节点可以代替旧雾节点继续为用户提供审计代理服务。PATCAR 被证明是满足证据不可伪造性、数据隐私保护、授权审计、密钥暴露抵制与审计者更换等安全属性，理论分析与实验结果进一步证实了 PATCAR 在密钥更新与标签更新阶段所引入的开销是少量的，并且本地端所需的计算开销是一个与文件划分块数无关的常量。

密钥暴露抵制：现有审计方案只考虑云服务器端的安全性，并不考虑用户端可能存在的安全问题：①用户可能出于经济考虑而选用一些廉价的软硬件密钥保护机制；②由于缺乏安全意识，用户容易从网上下载一些包含木马的软件；③当密钥较多时，用户可能由于密钥管理的复杂性而放弃对单个密钥的保护。以上因素都会导致签名密钥泄露。此外，密钥本身存在有效期，一旦过期，则需要立即进行更新。在现有云存储审计研究中，存在几个支持密钥更新的审计方案，但是它们或多或少存在一些缺陷，比如攻击者通过与撤销用户勾结可以推测出有效的签名密钥、外部审计者假定是完全可信的、繁重的证书开销影响审计效率等等。因此，设计一个安全高效的支持密钥更新的审计方案仍然是一个开放性问题。本文引入雾节点当作审计者。相比于一个受地域限制的审计实体，雾节点具有更强的移动性，处于用户与云服务器之间，代替用户与云服务器进行审计交互工作，降低了本地端的计算与通信开销。不同于传统审计模型中仅采用用户私钥为数据块生成标签，用户需要使用与雾节点共同协商的签名密钥来生成块标签。当用户或者雾节点密钥到期或者因为暴露原因需要进行密钥更新时，PATCAR 设计了一个标签与密钥更新机制，使得基于旧密钥的签名能有效地更新为新密钥的签名，并且计算量较大的标签更新操作由云服务器来完成，仅需要用户本地端执行少量的计算。

此外，云服务器不能获取签名密钥，也就是云服务器不能代表用户生成合法的块标签，保证了签名更新的安全性。

**审计者更换：**在现实场景中可能出现因更换审计者而导致审计服务中断的情况，这是因为审计者可能由于资源不足或身份过期等原因不能再为用户提供审计代理服务，或者审计者可能与云服务器勾结隐瞒数据损坏事实而失信于用户，此时需要及时选择新的审计者来代替旧审计者，保证审计服务的可持续性。然而，现有审计方案均假设在审计期间有且仅有一个审计者，因此不能适用于审计者更换场景下的云存储数据安全审计。因此，在实现密钥暴露抵制时考虑审计者更换需求也是非常必要的。在本文的审计过程中，当在职雾节点由于主动或者被动等原因需要离开审计任务时，一个新的雾节点能代替旧节点继续为用户提供数据审计服务，实现了审计者更换，同时离开的雾节点也不能获取新的签名密钥。此外，云服务器可以对审计者的身份进行认证，确保离开的雾节点不再能发起完整性挑战，也避免了外部攻击者冒充在职雾节点挑战云服务器，增强了审计安全性。

为同时实现密钥暴露抵制与审计者更换功能，本文利用双线性对与代理重签名思想构造了一种支持密钥更新与审计者更换的单副本审计方案 PATCAR，该方案能有效地抵制签名密钥暴露。不同于现有审计方案，PATCAR 的签名密钥并不是由用户的单独持有，而是由用户与雾节点的密钥以及时间密钥三部分计算而成，增强了密钥保护能力。同时针对签名密钥泄露、用户身份密钥到期、以及雾节点更换三种导致密钥更新的情况，该方案借鉴代理重签名思想分别设计了密钥与标签更新机制，将云服务器作为更新签名的中间代理，保证

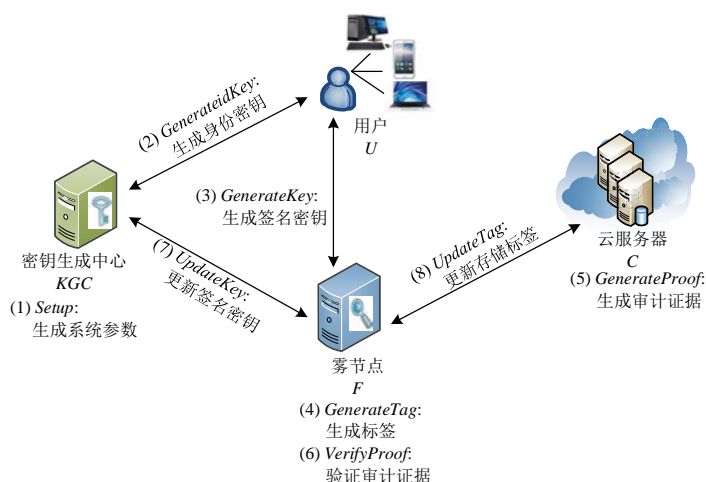


图 3: 方案 PATCAR 的系统模型

了基于旧密钥产生的签名可以有效地转化为基于新密钥的签名，有效地降低用户独自更新密钥与标签所需的开销。此外，基于所设计的密钥与标签更新机制，进一步实现了审计者可更换功能，保证了审计的可持续性。方案的主要思想归纳为：签名密钥由时间密钥以及雾节点所选秘密值组成。而时间密钥由用户身份密钥与系统主私钥计算而成。当使用签名密钥生成块标签时，时间密钥作为乘法因子参与计算。当签名密钥泄露时，通过更新时间密钥来



完成这个过程，并计算一个更新因子发送给云服务器进行标签更新。新标签等于将旧标签乘以更新因子来得到，根据标签的计算公式，将旧标签进行展开，再将旧时间密钥进行展开，然后将后两项进行合并计算，最后直接表示出新标签的计算形式，那这个过程证明了标签更新的正确性。也就是说，计算量较大的标签更新过程将有云服务器来执行。同时云服务器只能获取更新因子，无法获取新签名密钥。此外，安全分析证明 PATCAR 是可证明安全的，对比实验进一步证明了 PATCAR 在标签更新与审计者更换阶段仅引入了少量的计算与通信开销，并且本地端所消耗的计算开销是一个与数据块数无关的常量。

### 支持数据全动态与隐私保护的多副本审计方案 PDP-D

云存储中单副本完整性审计研究已经取得较为深入的进展，但是所提方案均不能适用于为具有重要价值的数​​据提供云存储服务，这是因为单副本存储仅选择在云端存储一个文件并删除本地副本。尽管云服务声称提供比本地更为可靠的服务，然而云上数据丢失与破坏事件频频发生。即使采用了单副本审计方案，一旦云服务器不能通过审计，则意味着云上存储的用户数据已经遭受破坏，而本地数据的删除导致数据难以恢复，最终给用户带来财产损失。因此，本文提出一个适用于具有重要价值的敏感数据存储的多副本公开审计方案 PDP-D，它通过采用在多个服务器上存储多个副本方式来提高数据的可用性和可恢复性。同时，为实现数据更新，设计了一种新的动态结构，使得动态更新所引入的通信开销与原始文件的数据块数无关，有效且安全地实现了多副本全动态更新。此外，采用随机掩蔽技术来产生一个区块的可分辨副本区块，避免了将内容暴露给未经授权的实体，达到数据隐私保护的​​目的。另外，在审计者与用户之间添加一个授权过程，避免了来自恶意或假装的审计者的恶意挑战。最后，形式化分析表明 PDP-D 是可证明安全的，并通过与现存多副本方案进行实验评估对比，证实了 PDP-D 在动态更新与审计阶段具有更小的计算与通信开销。

含敏感记录的重要数据存储需求：在云存储中，单副本存储模式只在云服务器上存储原始文件的一份副本，这种存储模式不适合具有重要价值的数​​据。这是因为如果云上仅存储一份数据，一旦出现设备故障或人为操作错误，就很可能造成存储数据被永久损坏或者丢失，从而给用户带来财产与名誉损失。比如云存储电子病历 (Electronic Medical Record, EMR) 数据包含患者身份、历史病历、社会保障号等私人信息，这些信息的损坏或者丢失将给用户造成不可估量的损失。多副本存储模式可以有效地避免数据损坏或丢失问题，它确保了只要一个存储副本没有被损坏，损坏的数据都可以被复原。与单副本存储模式相比，多副本存储有效地提升了数据的可用性与可恢复性。因此，采用多副本模式在云上存储高价值的重要数据是合理的选择。然而，一旦将数据上传到云端，用户就失去了对数据的直接控制权。同时云服务器可能会不小心删除或者篡改存储的用户数据。这时候，用户就需要检查云服务器是否真实存储了用户指定数目的副本，即多副本完整性审计。在多副本完整性审计中，数据隐私保护是需要考虑的重要需求。用户数据通常包含敏感信息，这些信息的泄露会​​给用户带来财产与名誉损失。为保护数据内容不会被泄露，以前的工作选择对上传数据进行加密，虽然解决



了隐私泄漏问题，但实际上只是将隐私保护归结于加密密钥的安全性，同时会带来额外的密钥管理成本。另一种隐私保护方法是在证据生成阶段掩盖信息，然而这种方法仅能保证隐私不会被暴露给审计者，无法避免云服务器获取审计隐私。所以，在多副本审计中，需要保护数据隐私，以确保除用户之外的未经授权的实体都不能获得数据明文。在数据处理阶段，PDP-D 采用随机掩蔽技术来生成原始文件中每个数据块的所有副本块，一方面保证每个副本块的可区分性，另一方面使得所有未授权实体无法获取数据明文。同时，PDP-D 将所有副本文件中位于同一索引的副本块的标签聚合成一个值，降低了标签上传与验证所需的计算与通信开销。

多副本数据全动态高效更新：用户需要对云存储数据进行更新，来确保数据的准确性与及时性。比如，在基于云的电子病历系统中，对于人为错误和医疗设备异常造成的不正确记录，需要及时清除，以免影响疾病的诊断。尽管在单副本研究中数据动态已经在某种程度上满足了用户的需要，然而它们仍然不适用于数据量比较大的情况，这是因为方案所带来的动态开销与数据块数成正比关系。目前也存在少量的多副本动态研究工作，然而它们在动态阶段引入的开销同时受到数据块数以及副本数的影响。因此，现有支持数据动态的完整性审计方案要么仅适用于单副本存储，要么存在更新效率低下的缺陷。因此，为高价值且包含敏感信息的数据设计一个多副本审计方案，在提供高效数据动态的同时实现针对未授权实体的隐私保护是一个值得研究的问题。本文设计了一个支持多副本批量更新的动态结构 M2HT。M2HT 将所有副本中具有相同索引的块视为一个叶子节点，确保了副本块更新的一致性，同时使得动态阶段消耗的通信开销与原始文件的划分块数无关，极大地提升了数据更新的效率。此外，为每个数据块附加时间属性，防止云服务器使用历史数据块欺骗用户。基于 M2HT，进一步提出一种高效的多副本审计方案 PDP-D，它能为具有重要价值且包含敏感信息的大数据提供多副本完整性审计服务，同时实现了数据完整性、可恢复性、动态更新、以及数据隐私保护。此外，PDP-D 在用户与审计者之间添加了一个授权过程，保证了只有通过用户授权的实体才可以挑战云服务器。

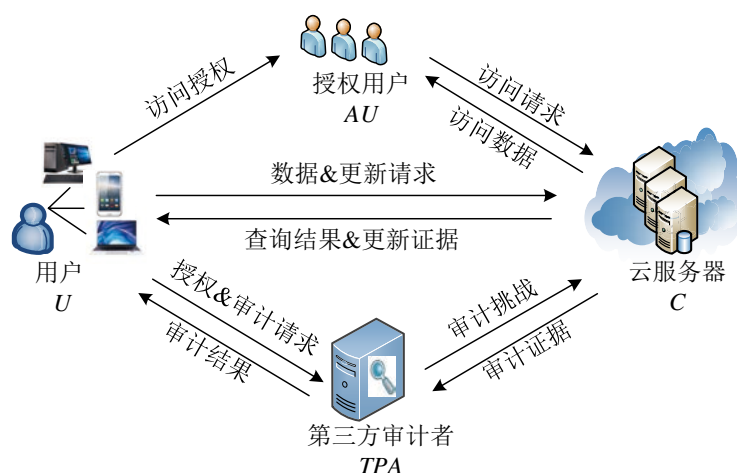


图 4: 方案 PDP-D 的系统模型

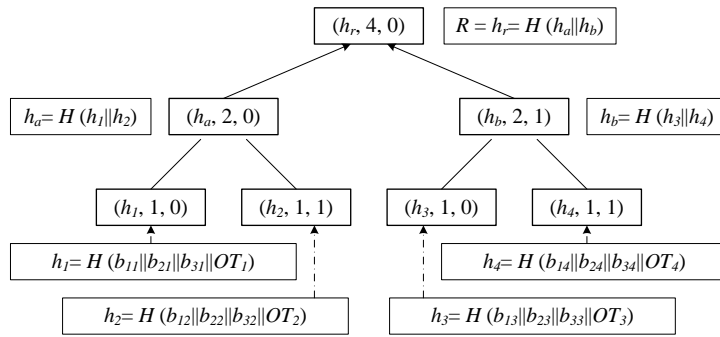


图 5: 动态结构 M2HT

针对多副本审计、数据隐私保护与多副本动态更新需求, 本文利用 BLS 签名与双线性技术提出了一个支持数据动态与隐私保护的多副本审计方案 PDP-D, 同时实现了数据完整性、可恢复性、动态更新、以及数据隐私保护。面对多副本动态开销与数据块数成正比关系的现状, 本文设计了一种基于 MHT 的新型动态结构 M2HT, 该结构的叶子结点存储了所有副本中具有相同索引的所有副本块的聚合哈希值, 并为树中的每个节点提供位置信息, 在提供数据块索引与内容验证的同时, 实现了动态所引入的通信开销与原始文件的划分块数无关。与其他多副本方案中所提出的 MHT 动态结构不同的是, M2HT 将所有副本中具有相同索引的副本块视为一个叶子节点。这种处理方式具有两个显著优点: ① M2HT 与 PDP-D 标签聚合方式一致, 即所有副本中具有相同索引的所有数据块只对应一个叶子节点与一个聚合标签。当用户需要更新原始文件中  $w$  个数据块时, 只需要更新 M2HT 中对应  $w$  个叶子节点, 而不是  $wm$  个叶子节点, 因此显著提高了多副本动态更新的效率。② M2HT 的高度比其他 MHT 结构的高度要低得多。例如, 当为包含  $n$  个块的文件创建  $m$  个副本时, 通过使用 M2HT, 将树的高度从原本的  $\log_2(nm)$  降为  $\log_2 n$ , 使得 M2HT 中的验证路径的长度与副本数无关, 有效地降低更新阶段所需的通信开销。此外, 为存储数据设置时间属性 OT 来记录创建或者更改时间, 防止云服务器使用未经修改的历史数据块来欺骗用户。基于该动态结构, 进一步提出了多副本审计方案 PDP-D, 在副本生成阶段将位于同一索引的块标签进行聚合, 并利用分区技术将数据块进行分区, 降低了审计过程所需的计算与通信开销。安全性分析表明 PDP-D 在随机预言机模式下是可证明安全的, 且满足授权审计、数据隐私保护与数据可恢复。通过与现有两个多副本动态方案进行对比, 证实了 PDP-D 在审计期间具有更低的通信和计算开销。

### 支持可变副本存储与公平审计的多副本审计方案 MDSS

大部分多副本方案与 PDP-D 都依赖于公钥基础设施, 在系统实现过程中需要消耗大量的证书管理开销。因此, 本文提出了一个基于无证书签名的支持可变副本存储与公平审计的多副本审计方案 MDSS, 它从根本上消除了昂贵的证书管理开销, 并避免了密钥托管问题。考虑目前多副本动态所引入的开销总是与副本数据呈现线性增长的趋势, MDSS 通过改进经典的 MHT 实现一个多副本动态结构 MD-MHT, 它不仅支持同时更新存储的多副本, 并且在动态阶

段所产生的通信开销与副本数无关。同时，为了满足灵活的存储需求，提出了可变副本数存储策略，允许用户确定将价值高的数据存储多个副本将价值低的数据存储少量副本，满足用户灵活的存储需求。此外，基于审计者或者云用户可能会诬陷云服务器的考虑，使用签名验证思想来实现公平审计。最后，通过可证明安全性分析与对比实现，证实了 MDSS 在随机预言模型中是安全的，且与所选的多副本动态审计方案相比，在计算复杂度与通信成本方面具有更高的效率。

多副本高效动态更新与可变副本存储需求：目前大部分多副本审计方案都建立在传统的公钥基础设施下，给审计过程带来了大量的证书管理开销。尽管存在几个应用基于身份的签名来减少证书开销的多副本方案，然而它们都没有考虑数据动态需求。在现有的少数几个多副本动态工作中，大部分工作为每个副本建立一个子树，导致动态开销是单副本动态开销的  $m$  倍，其中  $m$  为副本数。也就是，所构建的动态结构皆一个严重的效率缺陷：动态更新所需的开销始终与副本数成线性增长趋势，导致执行效率不高。此外，现有支持多副本更新的动态结构大多仅仅直接数据块的认证，存在云服务器可以使用其他有效数据块发动替换攻击的风险。因此，如何在实现数据块与序号双重验证的前提下，降低副本数对多副本动态开销的影响，是一个值得研究的问题。此外，现有的多副本审计方案始终假定每一个原始文件的所有数据块都默认被复制成相同份进行存储，然而实际上应该允许用户为价值低的数据块存储一个或几个副本，而为价值高的数据块存储多个副本，也就是，某份文件中的不同数据块应该被复制成不同份数进行存储，优化用户存储需求。可变副本数存储也为多副本动态更新的实现带来了挑战。本文设计了一种支持多副本更新的动态结构 MD-MHT。在 MD-MHT 中，每个节点同时存储哈希值与位置信息，实现了数据块内容与序号的双重验证，成功抵制了替换攻击。同时，MD-MHT 在叶子节点中存储同一数据块的所有副本块的聚合哈希值，实现了动态导致的通信开销与副本数无关，提升了多副本动态的效率。基于 MD-MHT，进一步提出了首个基于无证书签名的多副本审计方案 MDSS，既避免了证书管理开销，又解决了基于身份签名体制中存在的密钥托管问题。同时，提出一个可变副本数存储策略，允许用户在副本生成阶段自行确定同一文件中不同数据块的存储副本块数，优化了多副本存储的效率。

公平仲裁：以往的审计方案基于一个这样的模式：诚实的用户 / 审计者验可能不诚实的云服务器。但实际上，用户和审计者也可能是不诚实的。比如，即使云服务器通过验证后，用户在明知验证成功的前提下可以声称验证不成功，以便向云服务器索取赔偿。另一方面，审计者也可能与云服务器串通，向用户隐藏数据被破坏的事实。目前，仅有的几个允许检测审计者不诚实性的审计方案依靠保留审计过程中的所有随机挑战与对应证据来实现公平审计，给审计过程带来了大量的存储开销。在数据更新阶段，MDSS 采用签名交换确认思想要求用户与云服务器对每次的数据动态操作进行双向确认，体现为用户与云服务器相互验证各自对当前的根节点的签名。此外，引入一个可信的仲裁机构，它可以通过判定签名的有效性来解决参与实体的审计冲突纠纷，即实现了公平审计功能。

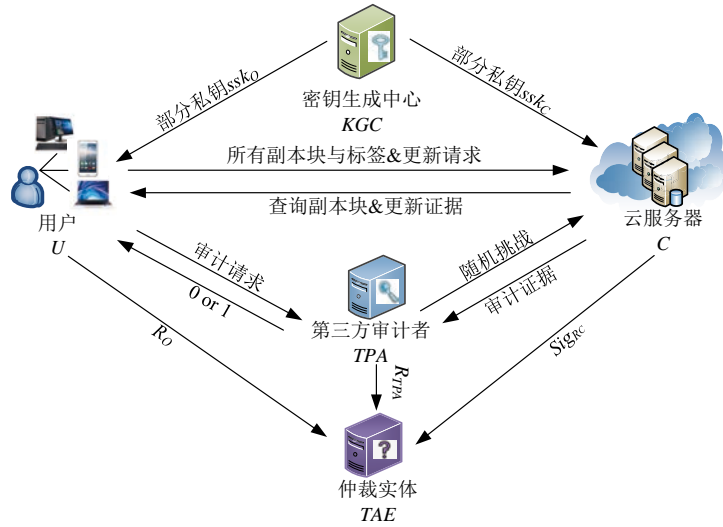


图 6: 方案 MDSS 的系统模型

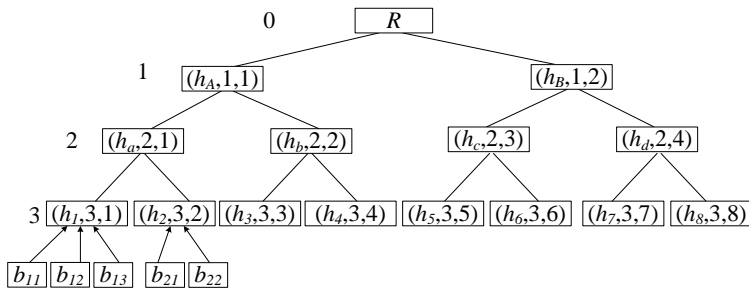


图 7: 动态结构 MD-MHT

为实现多副本高效动态更新与可变副本存储需求，并提供公平仲裁服务，本文结合无证书签名、双线性对与 MHT 结构，提出了一个支持可变副本存储与公平审计的多副本审计方案 MDSS，实现了数据动态、可变副本存储、与公平审计功能。针对多副本方案在动态阶段的开销与副本数成正比的问题，提出一个同时支持数据块内容与序号验证的动态结构 MD-MHT，通过在叶子结点中存储每个数据块的所有副本块的聚合哈希值，既可以为可变副本存储提供动态支持，又使得动态所需的通信开销与副本数无关，极大地提升了多副本更新的效率。在现有多副本方案中，动态开销与副本数  $m$  成  $O(m)/O(\log m)$  关系，并且这些方案不能用于可变副本存储。通过改造 MHT 结构，本文基于密码哈希函数  $H$  设计了一个支持多副本动态的数据结构 MD-MHT。在这个结构中，每个节点存储哈希值与节点的位置信息。为给每个节点确定唯一的位置信息，MD-MHT 从上到下标记节点的层次信息，从左到右标记节点的索引信息。按照从下到上的顺序，很容易得出所有节点的哈希值与位置信息，从而构建一个完整的哈希树。为了实现可变副本数审计以及支持动态批量更新，MD-MHT 设置每一个节点存储的哈希值是使用对应数据块的所有副本块计算而成，因此所引入的开销与副本数无关，并且能支持可变副本存储模式下的数据动态更新。基于 MD-MHT，本文利用无证书签名技术与可变副本存储策略设计了一个安全高效的多副本审计方案 MDSS，并通过引入一个可信的仲裁机构实现了审计的公平性。可证明安全性分

析证明了 MDSS 在随机预言模型下是安全的，且满足数据隐私保护与授权审计。此外，与现有多副本动态方案进行实验对比，MDSS 在完整性审计与数据更新阶段，具有更高的执行效率。

## 总结与展望

本文通过使用双线性对技术、无证书签名、雾计算、区块链等技术，构建了 4 个满足数据完整性、数据隐私保护、公开审计、数据动态、密钥暴露抵制、可变副本存储、审计公平性等不同安全需求的单副本与多副本审计方案，具体工作如下：

研究提出了一个适用于资源受限的用户本地端的单副本审计方案，包含标签计算外包以及公平审计方法。

研究提出了支持密钥更新与审计者更换的单副本审计方案，包含密钥暴露抵制策略与审计者更换机制。

研究提出了支持数据动态与隐私保护的多副本审计方案，包含数据隐私保护策略与支持数据块与序号双重认证的多副本动态结构。

研究提出了基于无证书签名的支持可变副本存储与公平审计的多副本审计方案（J.aa），包括支持可变副本存储的多副本高效动态结构以及公平仲裁机制。

### 作者简介：



第一作者：周磊，宿迁学院讲师，主要研究方向为主要研究方向为云存储安全。

E-mail:17149@squ.edu.cn

### 通信作者：



付安民，南京理工大学计算机科学与工程学院教授，博士生导师，主要研究方向为数据安全与隐私保护。

mail:fuam@njust.edu.cn



## 在“边缘智能”的科研道路上勤勉耕耘

### 个人简介

许小龙，男，汉族，1988年8月出生，共产党员，博士研究生学历，现任南京信息工程大学软件学院教授、博士生导师、软件工程专业负责人、服务计算与智能应用科研团队负责人。曾获江苏省计算机学会青年科技奖、江苏省高等学校科学技术研究成果二等奖（排名第1）、华东科技情报成果三等奖（排名第1）、2篇中国百篇最具影响国际学术论文（排名第1）、南京信息工程大学五四青年奖章、计算机领域知名国际会议IEEE UIC杰出领导奖、7篇国际会议最佳论文奖、2篇国际会议杰出论文奖等荣誉。连续两年入选全球高被引科学家（2022、2021年）；入选江苏省科协青年科技人才托举工程、江苏省科技副总。授权发明专利9项，实现落地转化3项。发表SCI检索论文100余篇，其中一区论文34篇，参与出版教材《计算机组成原理》（国家十二五规划教材、江苏省本科优秀培育教材）、气象信息交叉教材《Surfer气象制图指南》。指导中国国际“互联网+”大学生创新创业大赛全国银奖2项；中国大学生服务外包创新创业大赛一等奖1项、“中国软件杯”大学生软件设计大赛二等奖1项；江苏省优秀毕业论文奖1项。指导国家级大创项目2项；江苏省大创项目1项；江苏省研究生科研/实践创新计划项目4项。



### 坚持立德树人，培养高级创新人才

许小龙作为一名教育工作者，始终牢记“为党育人，为国育才”的教育使命，坚持以生为本，作为《计算机组成原理》的学校平台课负责人，积极钻研教学业务，参与撰写国家十二五规划教材1部。作为软件工程专业负责人，2021年带领本专业通过教育部工程教育专业认证。

Clarivate™



Highly Cited Researcher 2021



Clarivate™



Highly Cited Researcher 2021



图1 全球高被引学者

坚持科研育人，积极鼓励学生科研创新，培养学生开展创新研究工作的科学态度，指导国家级大学生创新训练项目2项，江苏省大学生创新训练项目1项，江苏省研究生科研/实践创新计划项目3项。作为指导老师所带领的团队获中国国际“互联网+”大学生创新创业大赛全国银奖2项、中国大学生服务外包创新创业大赛一等奖1项、中国软件杯大学生软件设计大赛二等奖1项，全国大学生算法设计与编程挑战赛金奖1项、全国信息融合挑战赛决赛二等奖1项、“北斗杯”全国青少年科技创新大赛全国三等奖1项。指导的学生中16人获得国家奖学金，团队毕业生相继至清华大学、中国科学院、南京大学、浙江大学、中国科学技术大学、复旦大学、东南大学等知名科研机构深造。相关育人成效被《南京日报》、光明网等主流媒体报道。



【南京日报】他们，被学生称为“大神”

2020-09-10 来源：南京日报 作者：谢智华 许天磊 严常坤 林雯 谈洁 设置



张绍铃（右）和学生

本科生跟他两年，人均一篇高水平论文



许小龙（前排中）和学生

带着9个本科生和3个硕士生做科研，不到两年的时间，学生们人手一篇高水平论文。南京信息工程大学计算机与软件学院入职3年的许小龙老师被学生们奉为“大神老师”，而他的团队则被送上了“信息创新先锋队”称号。

许小龙的研究方向是云计算、边缘计算、物联网、大数据等。听起来很热门的领域，许小龙却要求学生“能够耐得住寂寞、坐得住”。

“这些领域需要进行大量的外文文献阅读和高阶计算，学生需要具有良好的英文和数学功底，为他们‘宽、厚、实’的科研打下基础。”许小龙说，他要求学生和自己一样，忙着研究的时候，整天就是和文献、机器打交道，外面的喧嚣与自己无关，“能够耐得住寂寞，专心做好一件事。”

也因此，是否具有拼搏精神、能否耐得住寂寞，也成了他挑选学生的重点。今年刚刚跟着他读研的田昊来自一所普通的民办本科，“大学四年，他没有过得浑浑噩噩，而是在各方面尽力做到最好。这样的学生做科研一定差不到哪里去。”许小龙说。

为了解决大多数学生在科研选题、论文写作方面的普遍问题，许小龙会带着他们大量地阅读国外顶级期刊的最新论文，以此来拓宽学生的学术视野，激发学生的研究热情，同时也培养他们的文献分析能力和口头陈述能力，打下扎实的科研基本功。

“在许老师的引导和督促下，我们团队的学术氛围浓郁，同门师兄妹妹们都马不停蹄地看文献，做新选题。”学生方子介说，他大一期末时，通过文献的阅读发现，为车联网车载应用提供低延迟的服务方面，边缘计算有着非常好的表现。然而，由于不同车载应用所需的服务内容有一定的相似性，重复地处理相似甚至相同的服务内容，会导致边缘服务器中存储、计算和带宽资源的巨大浪费。在许小龙的指点下，他探索出了在车联网边缘计算环境下，通过交通流量预测实现计算卸载和内容缓存的方法。“在南京交通轨迹大数据上的实验结果证明，我提出的方法可以显著减少车载应用服务的延时和能耗。”方子介说，也正是这样的发现，让他的论文能够被顶级期刊录用。

图 2 育人成效被《南京日报》、光明网等主流媒体报道

## 强化基础研究，瞄准国际前沿技术

边缘智能在国家新一代人工智能发展规划中重要的战略性新兴产业，实现边缘“智能+”建设是国家产业信息化建设与行业数字化转型的重要工作。许小龙围绕边缘智能、服务计算等研究方向，依托国家自然科学基金重大研究计划、国家自然科学基金指南引导类 0-1 原创探索专项、国家自然科学基金项目重大研究计划、国家重点研发计划、兵团重点研究计划等项目开展研究工作。他带领项目团队，历经多年探索，揭示了行业大数据任务调度机理，率先研究基于大数据分析的任务调度理论，提出了兼顾隐私保护的安全可信的任务调度方法，建立了边云协同的任务调度技术，取得了系列突破性原创成果。他提出的边缘环境下基于区块链的安全计算卸载方法，在保障数据完整性和安全性的前提下，攻克了实现边缘基础设施的能耗、响应时延以及负载均衡的技术难题。相关成果分别入选 2020 年度与 2019 年度中国百篇最具影响国际学术论文。





图 3 中国百篇最具影响国际学术论文

作为一名青年科研工作者，他深知知行合一的重要性，积极参与国内外学术活动，任欧洲创新联盟会士，中国计算机学会服务计算专委会执行委员，江苏省计算机学会会员，Wireless Networks、Software: Practice and Experience 等国际期刊编委 / 客座编辑，IEEE UIC、IEEE HPCCC 等国际会议程序委员会主席 / 副主席，以及第十二届中国业务过程管理大会共同主席，尽自己所能引领正确科研价值观，促进国际学术合作和创新。



图 4 参与学术交流并作主题报告



### 推动成果落地，服务国家重大战略

许小龙积极投身我国首个国际大科学计划——深时数字地球，带领团队构建了全球首个基于大洋钻探计划的沉积物数据平台；并基于自然语言处理等技术，率先构建了自下而上与自上而下相融合的大洋沉积物知识图谱，同时实现了跨文献的表格信息融合与数据整合。贯彻落实党中央关于兵团向南发展的战略方针，带领南京信息工程大学、兵团石河子大学、新疆图木舒克市组成的技术攻关团队，深入新疆基层连队和社区调研，研发面向新疆师团的社区精细化管理系统，助推边疆地区高质量发展。依托学校大气科学特色，结合人工智能的技术优势，开展了气象雷达实时数据质量控制、基于静止轨道气象卫星的数据反演、短临回波外推等研究工作，为气象业务智能化、精准化、实时化的提供信息化技术支撑，服务国家气象信息化战略规划。

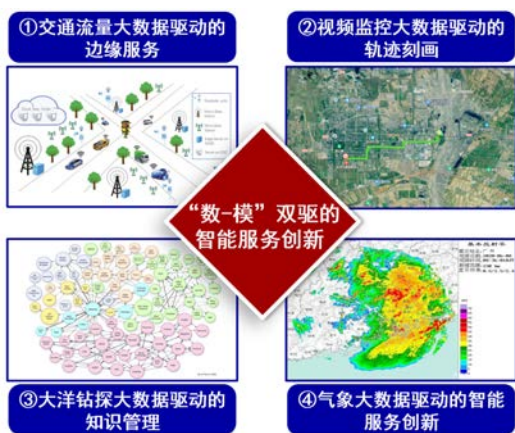


图 5 推动课题成果服务多个国家重大战略

## 学会动态

### 江苏省计算机学会获 2022 年度“科创江苏”专项行动计划结题认定

近日，江苏省科学技术协会下发了《关于认定“科创江苏”专项行动计划项目实施结果的通知》，对 2022 年度“科创江苏”专项行动计划项目中的“院士协同创新中心建设”、“会企合作”、“‘科创江苏’建设”三个类别 146 家项目单位项目结题予以认定，其中，我学会 3 个项目获 2022 年度“科创江苏”专项行动计划认定项目。

其中，与江苏安控智汇科技股份有限公司共建的“基于 RFID 的室内人员动作识别和定位系统研发”获“企业创新发展行动”项目单位认定，同时，我学会还获“省级学会科技服务站”、“专业科技服务团”项目学会认定。学会将严格按照“科创江苏”专项行动计划相关部署要求，发挥学会科技服务智能，为推动我省经济社会高质量发展贡献力量。

# 在“云端融合”的科研与实践道路上砥砺前行

——2023 年江苏省计算机学会青年科技奖东方教授

## 个人简介

东方，博士，东南大学青年首席教授，博士生导师，入选“国家万人计划青年拔尖人才”，现任东南大学大数据计算中心主任。同时担任 ACM 中国理事会常务理事、ACM 南京分会共同主席、江苏省计算机学会高性能计算专委会副主任。2011 年 3 月毕业于东南大学计算机科学与工程学院，获工学博士学位。博士毕业后留校任教，在江苏省“网络与信息安全”重点实验室从事教学科研工作，主要研究方向为云计算、物联网与边缘计算。作为项目负责人主持科技创新 2030 重大项目课题、国家重点研发计划项目子课题以及国家自然科学基金面上项目等多项国家级项目。参加了丁肇中教授领导的 AMS 大型物理实验，建设完成东南大学云计算中心及东南大学 AMS 科学数据处理中心（AMS-02 SOC）。在云计算、物联网与边缘计算等领域共发表论文 100 余篇，包括国际一流学术会议 IEEE INFOCOM、WWW、IEEE ICNP、IEEE ICDCS，以及国际一流学术期刊 ACM/IEEE TON、IEEE TMC、IEEE TSC 等。研究成果荣获由教育部主办的全国高校云计算应用创新大赛一等奖，并在南钢集团等工业互联网领域以及 AMS 实验等科学计算领域进行了广泛应用，取得了显著的经济社会效益。





## 坚守独立探索精神，以高水平科研推动学科发展

东方教授在云计算、物联网与工业互联网领域作出重要学术贡献，入选“国家万人计划青年拔尖人才”。近年来以“复杂场景端边云高效协同传算”为核心科学问题，针对端-云传输延时高与云侧计算效率低的问题，分别从“端边云跨协议网络传输优化”、“算网融合协同计算优化”两方面展开了深入的理论研究与技术研发，取得了一系列研究成果，并研发了高可用端边云协同计算支撑平台。发表高水平学术论文 100 余篇，授权发明专利 15 项，获批团体标准一项。以上工作得到科技部与自然科学基金委的大力支持，先后承担科技创新 2030 新一代人工智能重大项目课题“感知智能驱动的端边云融合计算”并以项目技术负责人身份承担国家重点研发计划项目“面向工业互联网的智能云端协作关键技术及系统”。相关成果为实现物联网数据高效分析提供了重要的关键技术与平台支撑。

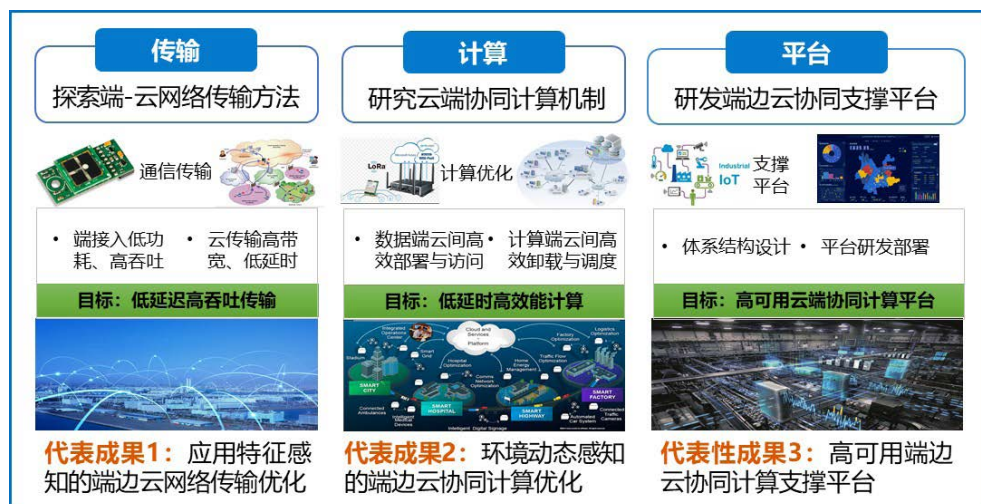


图 1 面向复杂场景的端边云高效协同传算

## 发挥科创引领作用，以示范性应用促进社会发展

东方教授积极促进研究成果的应用示范。为响应国家制造强国战略需求，将成果应用于钢铁制造行业，在南钢集团滚期排产、能源管理、安全保障领域进行应用落地，重点解决钢铁制造过程中由于数据传输与处理不及时导致的生产效率低、能源消耗高、安全隐患大等问题，实现钢板成材率提升、设备故障率降低、煤气放散率下降，累计新增产值超 2.7 亿元。并受邀参加第二十二届中国国际工业博览会，得到了业界和同行的广泛关注。在空间探测科学大数据处理领域，将成果应用于诺贝尔奖获得者丁肇中教授领导的 AMS 物理实验，实现科学数据的端云低延时传输、高效能计算，将 PB 级数据的查询处理由原来的分钟级降低到秒级，助力东南大学成为世界范围内 AMS 实验最主要的数据处理中心之一，为 AMS 实验取得阶段性成果做出了极为重要的贡献，也为我国开展大型空间科学实验提供了重要参考。

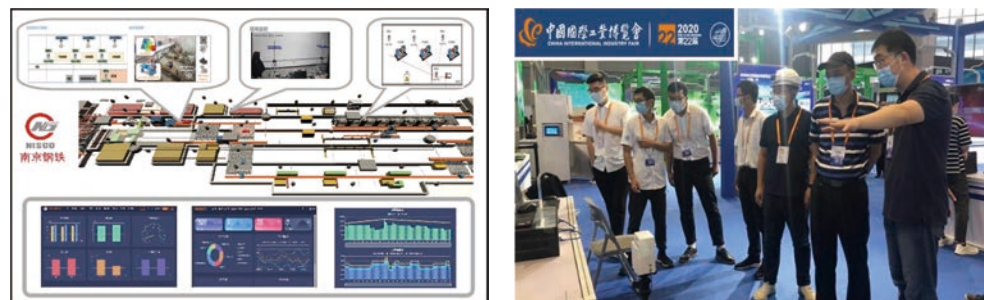


图 2 端边云高效协同研究成果在工业制造领域的应用

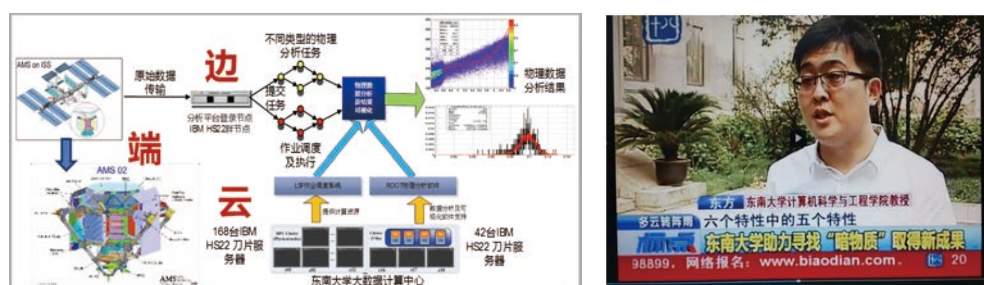


图 3 端边云高效协同研究成果在 AMS 科学实验数据分析领域的应用

### 落实立德树人任务，扎根国家教育培养创新人才

东方教授任教以来，始终坚持以知识传授、能力培养、德行树立、价值创造为育人理念。在科研育人方面，近年来组建了ACM南京学生分会（中国大陆首个学生分会）以及东南大学-IBM云计算联合研究中心，围绕云计算相关技术开展联合研究与人才培养，共指导（含联合指导）50多名博士/硕士研究生，其中多名博士生获评“ACM南京分会优博奖”（代表江苏范围，每年仅3名）。基于以上工作成果，东方教授获评教育部-IBM高校合作项目年度优秀教师、东南大学“三育人”先进个人，并连续三年获得东南大学吾爱吾师计算机、软件学院最受欢迎的老师称号。在实践育人方面，自2015年起连续四年牵头组织了国家级赛事“全国高校云计算应用创新大赛”，该赛事累计参赛人数超万人，取得了较大的社会影响力，成为云计算领域全国范围内的重要旗舰赛事。与此同时，东方教授自2013年起依托江苏省计算机学会云计算专委会以及ACM南京分会，成功组织了十届云计算与大数据国际会议（CBD 2013-2022），为广大老师和学生提供了良好的学习交流的平台，为云计算、物联网等相关领域的科研发展、人才培养以及学术平台的建设持续做出贡献。



# 啥?! 能打败 GPT4 的“强人工智能”来了?

最近 OpenAI 发布的 ChatGPT 和 GPT-4 大型语言模型，几乎引爆了全民对人工智能话题讨论的“火药桶”，相信不少人的朋友圈已经被带“GPT”的字条给刷屏了。

笔者身边的朋友们也分了两派，激进派认为迈向强人工智能的奇点已经到来，未来数据量和算力规模再涨一波后，它就能替代掉绝大多数文字工作者了，随后所有没创造力的工种都将面临失业；保守派则认为它只是一个顶尖的生产力工具，但无法掌握意识、感情、创造力等人类与生俱来的能力，难以成为强人工智能。

而笔者属于“好吃懒做”派，选择直接问 ChatGPT，以下是它的回答：

**yu** ChatGPT，未来不断优化下你能成为强人工智能吗

作为一个 AI 模型，我的训练和性能会随着时间和技术的进步而不断优化。然而，要成为强人工智能，我需要比当前的技术水平更进一步地发展。

强人工智能需要具备一定的智能、创造力、自我意识等人类所具有的高级认知能力。虽然当前的 AI 技术已经取得了很大的进展，但要实现这样的目标还需要克服很多技术难题和伦理问题。

目前的 AI 模型，如我所属的 GPT 系列，主要是基于预测和模式识别，缺乏创造性和自主性。而强人工智能需要能够进行自我学习和创造，并具有一定的情感和意识，这些目前还远远无法实现。

因此，虽然我会不断优化和进化，但要成为强人工智能还需要时间和技术的不断突破。

图片来源：截图自 ChatGPT

强人工智能能否实现目前尚未可知，但毫无疑问，目前人脑还是整个地球上最智能的复杂系统。

可能有人会问：我们为什么不直接用生物的神经元作为网络基本单元，从头复刻一个和人脑类似的智能系统呢？这会不会是实现强人工智能的捷径呢？好问题，欢迎来到本文的主题——“类器官智能（Organoid Intelligence, OI）”。


## 01

### 碳基和硅基并肩作战

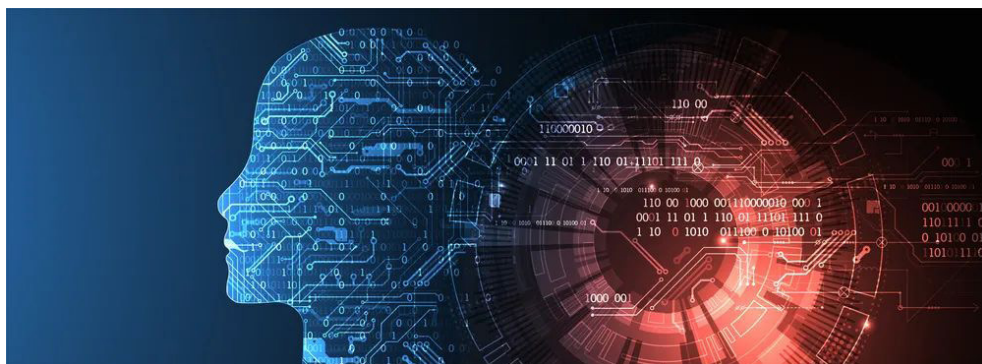
#### 就问你怕不怕！

类器官智能一词是今年2月28日，由约翰斯·霍普金斯大学的托马斯·哈东（Thomas Hartung）团队提出来的新概念。

简单讲，类器官就是把干细胞在体外三维环境中培养成有一定结构和功能的组织类似物，目前已成功构建包括小肠、胃、结肠、膀胱、肝脏、心脏、胰脏、肾脏、大脑等多种器官组织。

或许在不久的将来，你身体某个部分出了问题，都可以用类器官进行修复或替代，比如熬夜搬砖爆肝了换个肝  (bushi)。

而类器官智能就是将体外培养的大脑器官组织作为生物硬件，通过与外部的电子设备相连来实现生物计算，我们人工构建的大脑，就是整台计算机的CPU+GPU，想想是不是很科幻呢？



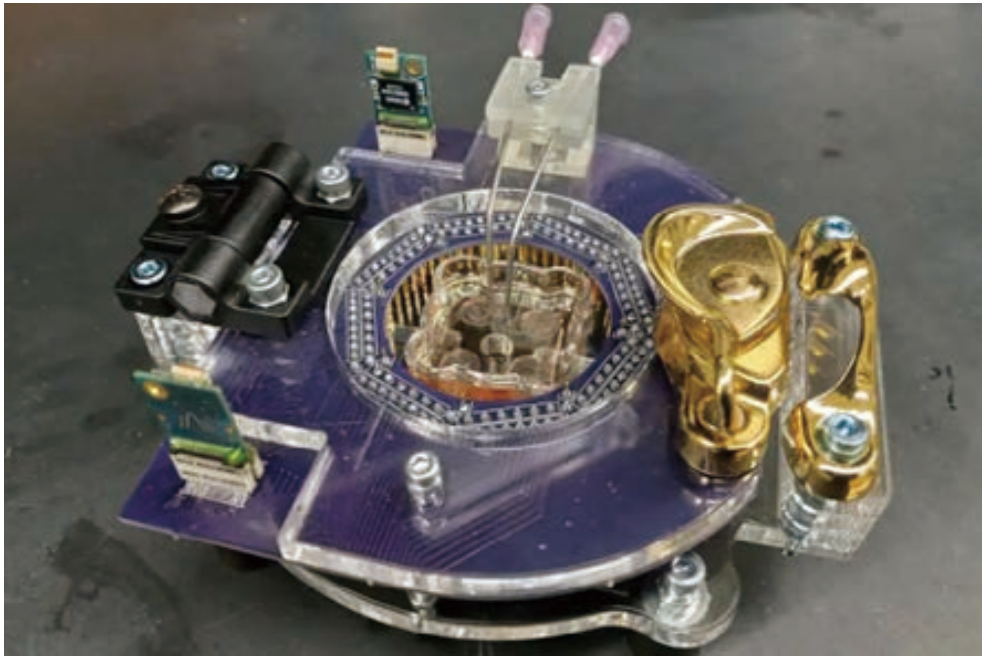
实际上，目前很多研究已经证明了该路径的可行性，就拿最近的新闻来说，伊利诺伊大学厄巴纳-香槟分校的 Andrew Dou 团队就培育了 8 万多个通过重编程小鼠干细胞得到的神经元，并将其放置在光纤和电极网格之间，接受 10 种不同模式类型的电脉冲和光信号刺激，这些元件都被放在一个恒温的箱子中，来保持神经元的活性。

经过一小时训练后，研究人员发现这群神经元每次在接受模式相同的刺激时，发出的信号也相同。研究人员还用 F1 分数来量化该神经网络模式识别的效率。简单来说，0 最差 1 最好，最后测得该活体神经网络搭建的计算机 F1 得分为 0.98，可以算相当智能了。



除模式分类任务外，研究人员还把“盒中之脑”连接到用活体肌肉组织制造的机器人中，使大脑能够通过肌肉感知周围的环境变化并处理这些信息。

将活细胞和储层计算技术相结合，能使神经元和芯片，或者说碳基和硅基的智能单元，协作完成信号识别和处理任务，这就是碳基生物和硅基生物智能合一的原初形态，虽然目前长得还比较丑。



中间放生物神经元的类器官智能机器人。图片来源：参考文献 [3]

再往前推也有脑洞大开的研究。

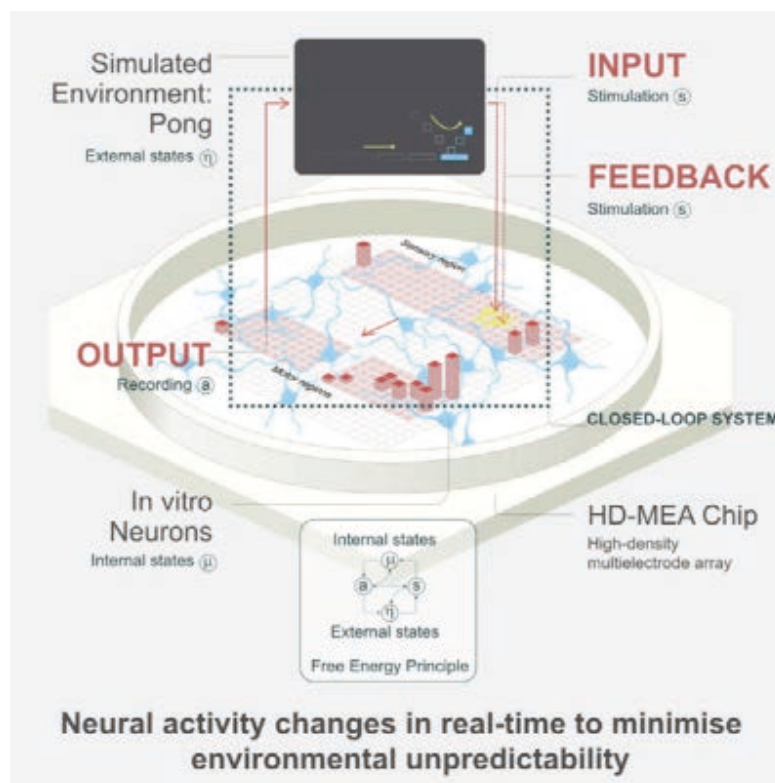
例如去年 12 月，澳大利亚生物科技初创公司 Cortical Labs 就用人类脑细胞培育的“碟中大脑”（Dish Brain）学会了打乒乓球。

研究团队将从人类干细胞诱导分化而来的人类神经元，与高密度多电极阵列与计算机集成，使“乒乓球”游戏中的电信号传到微电极阵列上，进而告诉神经元“乒乓球”的位置，神经元再通过彼此间电信号的交流达成一致后，控制“球拍”的移动，从而将“乒乓球”打回去。

神奇的是，这群“碟中大脑”仅用 5 分钟就学会了这款游戏，而相似规模的人工神经网络可能要 90 分钟左右才能学会。

尽管该论文的第一作者坚称“碟中大脑”已经将自己代入到游戏中的球拍了，但类器官智能是否能像人类一样“自主而有意识”地摄取、分辨并回应外界信息，仍是个值得探讨的问题。





会打乒乓球的“碟中大脑”。图片来源：参考文献 [4]

## 02

### 误打误撞出的类器官智能技术

大脑的类器官技术起初并不是为了实现“类器官智能”。自2019年相关研究取得突破以来，大部分的工作都围绕研究大脑的发育和疾病，或是修复大脑的残缺部分为主。

比如去年10月《自然（Nature）》主刊上的论文显示，科学家首次将人脑神经元移植到大鼠脑中并形成连接，从而控制大鼠行为。

4个月后，《细胞（Cell）》子刊上的研究又显示人的类脑器官植入到大鼠脑部后不仅实现了有效连接，还能对视觉刺激做出反应，说明它已经整合到大脑中发挥功能了，这个成果可以作为一种恢复皮质功能的治疗策略。

然而总有些充满创意的科学家喜欢搭建出一些让人眼前一亮的科技网络，比如这个类器官智能技术。

当然这个技术目前还是襁褓中的婴儿，等它渐渐成熟并与当前基于深度学习人工智能技术进行比较，究竟是变成科研领域中永远“未来可期”的空中楼阁，还是能成为新一波类脑智能浪潮中的弄潮儿，直奔强人工智能而去呢？让我们拭目以待。

类器官智能作为由生物智能单元——神经元组成的网络，它的优势集中体现在以下几方面：



### 1. 能耗低

一条斑马鱼幼虫在水中，成功捕食猎物并从天敌眼皮底下逃脱的大脑功耗仅为 1 微瓦，成年人类的大脑功耗仅为 20 瓦，而目前依赖深度学习的服务器集群的功耗通常在 1 百万瓦左右，拥有最强算力的 Frontier 超算功耗将近 21 兆瓦，听起来并不是很环保。

### 2. 少样本学习

生物通常能使用较少的观察活动来进行学习。人类使用大约 10 个训练样本就能完成一个简单的“物体类型相同或不同”的任务，蜜蜂这类昆虫也只需要 100 个训练样本就能学会，而这个训练数据量在深度神经网络中往往分类效果不佳。

AlphaGo 系统接受了 16 万场围棋比赛的训练数据，每天训练 5 小时的棋手则需要风雨无阻地连续下 175 年的棋，才能打完这么多场比赛，可见大脑在学习活动中不需太多数据也能达到很高的训练效率，同时储存能力惊人（大约为 2500 兆字节）。

### 3. 提供优秀的脑机接口界面

将电脑或电极输出的电信号传输给类器官脑组织，而植入人脑后的类器官又能充分整合到脑组织内执行其功能，这种界面具有很好的生理性，可将接口对大脑的损害几乎降到最低，同时又能将机器智能与自然智能一体化，从而实现新的智能形式。

当然，如何将电极连接到小型和三维的类器官上也是一个需要考虑的问题，这更需要高质量的 3D 电极作为支撑。如果以后要构建类人机器人，从生理学上来看，使用神经元与模拟肌肉组织相连也会更为便捷。

## 03

### 残忍还是先进，这是个问题

二十一世纪既是生物技术的世纪，也是信息技术的世纪，更是交叉科学的世纪。

目前，多支研究团队使用基因编辑技术和光遗传学技术构建了具有特定功能的神经网络，并通过纳米技术和生物打印技术来构建更加复杂的类器官框架。可以预料到，以后培养出来的脑组织将会有更精细的结构与更特异性的功能。

“类器官智能”的初衷还是想利用生物计算更快、更高效且低能耗的优势来构建活体计算机，从而在较为复杂的任务上能实现比传统硅基计算机更优秀的表现，并且通过电脉冲信号能向计算机芯片发送或接收指令，实现碳基和硅基神经网络的协调计算，融合两者的相对优势，创造出更加智能的计算系统。

**但目前类器官智能的挑战主要分两块：伦理和技术。**

首先是伦理问题。

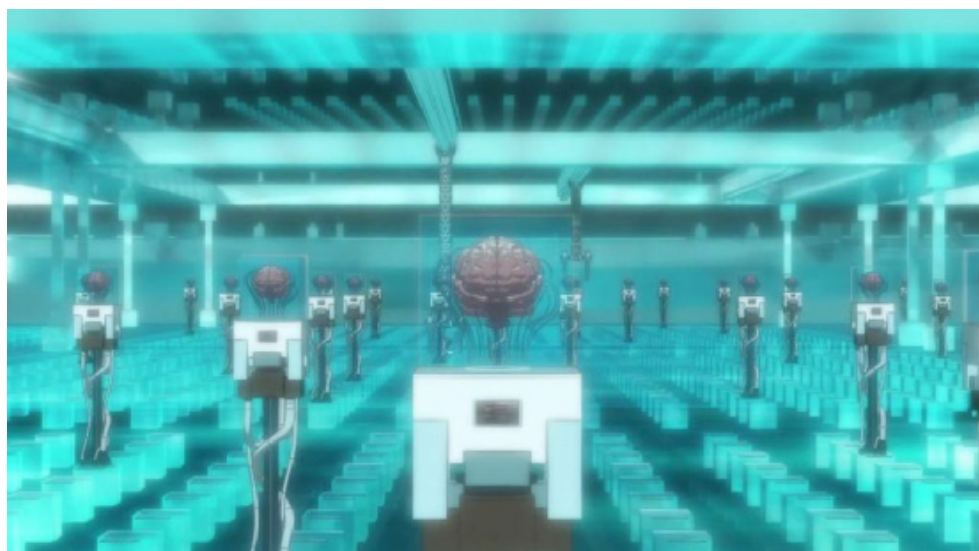
培养 10 周后的类器官会显示出髓鞘形成等怀孕 20 周胎儿的特征，而且带有信息输入的刺激会影响类器官发育，结构上变得更为复杂。接受输入，产生输出，与周围的环境信息产生互动，并建立原始记忆，它是否会有意识？外界给予的电信号刺激输入又是否会引起这些

大脑的“痛苦”，这些是伦理机构一直在密切关注的问题。

找到意识产生的充要生理条件是神经科学最困难的问题之一，目前揭示意识神经基础的工作也在进行中，这将为类器官伦理条例提供很好的参考建议。类器官智能本身也不是为了重建人类意识，而是提供生物学习，认知与计算的功能基础。

如果强人工智能的实现需要大量神经元作为基础的话，这种智能又和我们人类智能有何不同？极端一点，我们直接把尚未形成自主意识的胚胎神经管切下来（这部分以后会形成人脑），然后放入培养箱中培养成一个具有 860 亿节点体量的网络系统，并将其与计算机芯片或服务器相连，处理各种任务，这种强人工智能的实现是否人道呢？

笔者相信这项技术即便研发成功，也会因相当复杂的伦理问题而被法律法规所禁止，就像应用于人类的克隆技术和基因编辑技术一样。



动漫《心理测量者》中的强人工智能——西比拉系统。图片来源：动漫《心理测量者》

其次是技术问题。

前文提及的托马斯·哈东团队使用的单个类脑器官大约包含 5 万个神经元，安德鲁团队则是使用了 8 万个。麻雀虽小，但五脏俱全，当前培养的类脑器官能再现大脑的组织结构和功能，并有覆盖髓鞘的轴突，自发的电生理活动，复杂的振荡行为，高细胞密度和分层模式，甚至还有少突胶质细胞、小胶质细胞和星形胶质细胞等多种细胞类型。

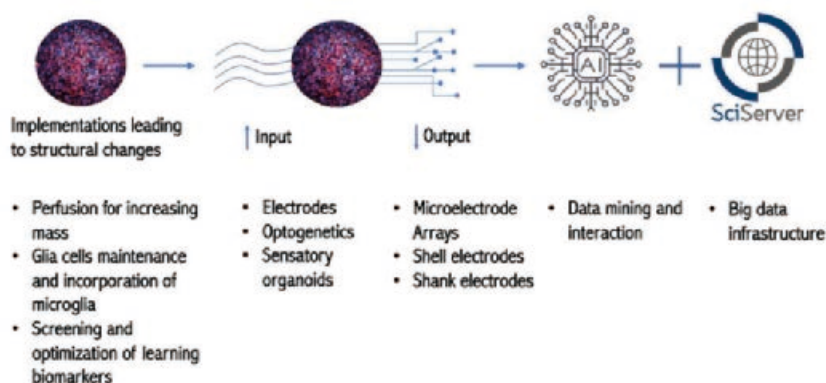
但问题是，如何使这些类器官产生学习活动，并对其计算能力加以利用？而且，多大规模的神经元数目才能表现出高智能？下一阶段研究人员将基于现有的分化方案扩大培养规模，构建 1000 万个神经细胞的活体计算机，至于表现出来的智能水平如何，让我们拭目以待。

与此同时，如何对神经元间的连接做出调整，使其能更好地实现功能也需要考虑。尽管大脑类器官可能实现分子特征的时空特征，但无法反映人脑脑区的拓扑结构和神经元环路的复杂性及特异性，而这些可能恰恰是实现大脑高级功能的基础。如何对大量杂乱的神经元连

接进行合理的布线，用分子信号诱导相关功能环路的产生，也是未来制作特定功能导向型的生物神经网络所要考虑的问题。

除此之外，目前仍无法高效完整地完对人脑类器官的信号输入与输出记录的记录，研究人员正在致力于研发专用于类脑器官的 3D 脑机接口以及相应探针，例如 3D 微电极阵列（MEA）、神经像素硅探针等，完成与外界信息端口的精准对接。这些技术的推进能够在一定程度上解决这方面的难题。

总的来说，尽管当前人工智能远不如人脑学习能力全面且高效，只能在经过大规模预训练的任务上表现出色，但类器官智能还难以复现最简单的大脑，想要“战胜”以 GPT 系列打头阵的硅基生物，恐怕还有很长一段路要走，尤其使用的生物神经元数目越多，所面临的伦理挑战就越大。对智能这条路探索的尽头，很可能碳基和硅基两种智能就会走到一块儿去。

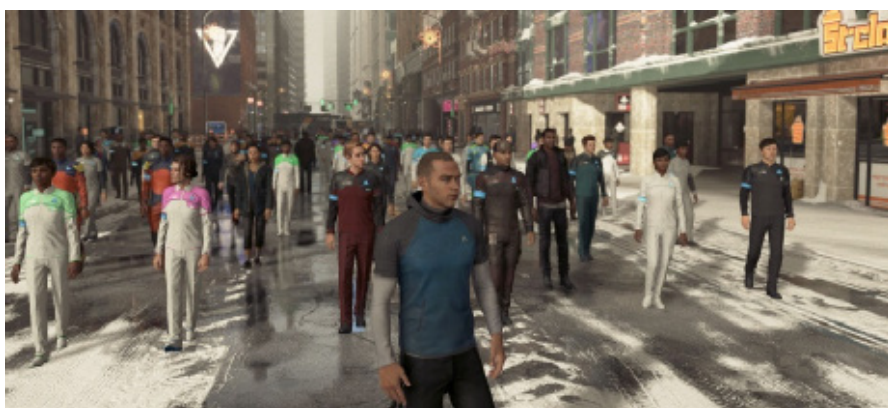


实现类器官智能的路线图。图片来源：参考文献 [2]

在实现强人工智能前，我们不妨再问问自己，为什么要实现强人工智能？

如果仅仅是需要服从听话且好用的工具，那么只需训练出在特定任务执行出色的智能神经网络即可，意识、情感、创造力这些都是不稳定因素，对于任何低技术需求且具有高重复性的工作来讲根本没必要，而且还会导致社会的管理成本上升。

这么说来，可能对人类来说，像 ChatGPT 一类的模型已经是最理想的智能系统了，希望看到这篇文章的你们在新一波的人工智能浪潮中都能用好这个工具，最大化地提高生产力。



人类和仿生人。图片来源：游戏《底特律：变人》

# 审计大数据智能分析关键技术与应用

——2022年江苏省计算机学会科学技术奖一等奖

## 基本情况

项目名称：审计大数据智能分析关键技术与应用

完成人：徐超、王海、徐泽水、钱钢、黄佳佳、陈勇

完成单位：南京审计大学、四川大学

## 项目简介

审计是维护国家安全和经济社会健康运行的“免疫系统”，是党和国家监督体系的重要组成部分。习近平在中央审计委员会第一次会议上指出，要优化审计资源配置，做到应审尽审，努力构建集中统一、全面覆盖、权威高效的审计监督体系。侯凯审计长指出，深入开展研究型审计是实现总书记宏伟目标的必由之路，审计行业和科研人员必须坚持“新兴技术赋能审计”的策略。研究型审计视角下，审计全覆盖，不仅是审计内容、目标的全覆盖，更是对审计数据的全覆盖。传统审计技术在实现审计数据全覆盖时至少存在以下困难：（1）以分析财务数据为主，缺乏对文本类型的不确定信息的有效处理技术；（2）无法有效融合具有巨大潜在价值的、呈现多种数据类型的业务数据、行业数据及跨行业跨领域数据；（3）缺乏对海量数据的分析能力，无法对审计对象发表全面的审计意见，从而无法实现审计内容的全覆盖。习近平总书记在中央审计委员会第一次会议上强调，要坚持科技强审，加强审计信息化建设，要把自然语义分析、智能关系发现、多源数据关联等数据分析方法引入审计业务中，降低大数据应用的技术门槛，破解快速实施跨行业、跨部门、跨地区大数据关联分析的难题。

### 项目突出性的创新工作：

本项目依托多个纵向国家基金项目与横向审计实务项目，为《2016 研究前沿》及《2018 研究前沿》报告中的“犹豫模糊集理论及其在决策中的应用”、“群体决策一致性的模糊语



言建模”、“大数据背景下的管理学问题研究”等研究前沿做了大量奠基性工作，在不确定型审计大数据处理的 3 个方面取得了一系列原创性或创新性研究成果，具体包括：

(1) 系统地提出了虚拟语言标度的语法、语义模型和计算模型，为科学处理包含主观不确定的审计评估数据以及具有不确定语义的审计指标数据和审计文本数据提供了理论基础。

(2) 针对包含不确定语义的审计数据和包含主观不确定（定性）的审计评估数据，在项目成果（1）的基础上，创新性地提出了概率语言集这一新模型，并提出了基于该模型的审计数据融合技术，以及审计多属性综合评估方法，扩展了审计评估技术。

(3) 针对残缺的定量型审计数据，提出了基于犹豫模糊理论的数据融合方法、分类算法和智能决策算法，既实现了不确定数据的融合，扩大了审计数据取证范围，也创新了适用于非平衡审计数据分析的人工智能技术。

本项目提出的关键技术不仅丰富了决策分析、人工智能研究成果，且极大地拓展了可用于审计分析的数据类型，向实施全覆盖审计迈出了坚实的一步，为科技强审提供了科学高效的技术支撑。本项目技术已经在领导干部自然资源资产离任审计评价体系研究、南京市鼓楼区资源环境审计研究等审计项目中取得了成功的应用；理论成果与应用成果先后被江苏省审计厅、南京市公安局、南京市政务服务管理办公室、珠海市审计局采纳并应用推广。

本项目代表性论文共计被引超过 940 次，其中他引超过 750 次。项目成果受到了中国工程院院士、审计长、中国审计学会会长、全国政协委员等的积极评价；受到中国工程院院士、杰青、优青、长江学者、国际权威期刊主编、全球高被引科学家、IEEE 终身 Fellow、国际模糊集协会 Fellow、英国工程技术协会 (IET) Fellow 等来自中、美、英、日、加等数十个国家的学者的高度关注和广泛肯定。

### 主要创新技术：

本项目依托多个纵向国家基金项目与横向审计实务项目，历时多年，在不确定审计数据处理方面取得了一系列原创性或创新性研究成果。主要创新点及研究内容如图 1 所示。



图 1 审计大数据智能分析关键技术主要创新点及研究内容

**主要创新点 1:** 系统地提出了虚拟语言标度的语法、语义模型和计算模型，为科学处理包含主观不确定的审计评估数据以及具有不确定语义的审计指标数据和审计文本数据提供了理论基础。

具有不确定性的信息广泛存在于审计数据中。传统审计技术简单地用整数表示对应的语言标度，忽略了其中的不确定性，降低了审计结论的科学性和可靠性。项目完成人基于模糊数学、词语计算等基础理论，系统建立了虚拟语言标度的语义、语法和推理模型，为具有语义不确定的信息表示与融合奠定了基础，也为词语计算提供了新的理论工具。本项创新点已成为相关领域研究的必备参考文献之一，其主要成果如下：

(1) 完整构建了虚拟语言标度的表示模型。以模糊集和传统语言变量为基础，在确定原始标度的语义前提下，用模糊映射刻画语义上的不确定程度；语法上，具有不确定性的虚拟语言标度可通过原始标度和一个常数构造；语义上，根据原始标度的语义类型，虚拟语言标度的不确定语义由原始标度语义和模糊映射确定。

(2) 系统构建了基于虚拟语言标度的计算与推理模型。基于上述语义模型，首先构造了全体虚拟语言标度上的全序关系，并提出了虚拟语言标度的否算子，表示逻辑上的否定；基于语义和语法模型，提出了基于虚拟语言标度下标的信息融合理论框架，并构造了具有良好性质的信息融合算子与不确定推理方法。

上述表示模型与推理模型兼具科学性、易用性、可解释性和直觉上的合理性，被广泛应用于基于语义信息的决策分析研究中。本项成果丰富了不确定决策领域的信息表示与推理理论和方法。成果完成人在本项创新点上发表了论文 3 篇，其中 ESI 高被引论文 1 篇，Web of Science 中被引次 132 次（他引 63 次），出版学术专著 1 部。

**主要创新点 2:** 针对包含不确定语义的审计数据和包含主观不确定（定性）的审计评估数据，在项目创新点一的基础上，创新性地提出了概率语言集这一新模型，并提出了基于该模型的审计数据融合技术，以及审计多属性综合评估方法，扩展了审计评估技术。

随着决策环境的日益复杂和审计对象涉及指标的日益增多，如何在审计评估或决策分析中考虑复杂的评价信息，挖掘不确定、不一致信息中的真实值，是一个重要的科学问题。为避免传统审计技术面对此类问题的不足，项目完成人创新性的提出了概率语言集。自提出以来，该成果迅速引起国内外学者的高度关注，并在公司战略管理、推荐系统、客户关系管理等领域得到了成功的应用。项目完成人在此领域始终引领国际研究前沿，取得了若干原创性研究成果，主要体现在以下几个方面：

(1) 创建了概率语言集的基础理论。在虚拟语言标度模型的基础了，提出了概率语言集，它是若干虚拟语言标度的一个部分或完全的概率分布，扩展了语言变量的取值范围，以适应复杂不确定环境；全面定义了概率语言集的规范化方法、序关系理论、不确定度量方式和基本运算法则，并保证了良好的数学性质。理论上，概率语言集为现有研究中所有不确定语言表达式提供了统一的表达和计算范式。

(2) 建立了基于概率语言集的不确定信息融合理论。信息融合作为决策分析的研究基础，



在多属性决策（如审计评估）中占据着重要地位。基于概率语言集上的偏序关系和基本运算，先后提出了适用于不同融合策略的信息融合算子和融合方法。针对基数和序数类审计指标，提出了概率语言加权算术平均和几何平均算子；针对有相互关联关系的审计指标，提出了概率语言 Choquet 积分算子和 DEMATEL 算子。

(3) 提出了若干基于概率语言集的多属性评估方法。为科学利用群体（多个数据源或多名专家）不确定信息，做出合理的审计评估结论，提出了基于概率语言集的多属性决策或多属性评估方法，包括基于距离测度和理想点（基期标准或行业水平）的 TOPSIS 方法、基于指标优先级的 PROMETHEE 和 ELECTRE 多属性决策方法等。特别地，为促进领导干部形成积极的发展观，提出了适用于自然资源资产经济责任审计的动态权重综合评估方法，避免领导干部过度关注少数权重大的指标，造成自然资源的损失或浪费。

(4) 创建了基于概率语言集的偏好分析理论与方法。偏好分析是解决复杂评估问题的指标权重的基础。针对自然资源资产离任审计等复杂评估体系，提出了基于概率语言集的偏好分析理论与指标权重获取方法。相比传统赋权方法，该方法以概率和模糊不确定的方式表示主观不确定评估信息，以随机分析为技术手段获取可信的指标权重，并以定量地方式度量不确定程度，为审计风险提供监测依据；该方法在专家评估信息不一致的情形有稳定、可靠的输出，并且输出具有直观可解释性，易于在审计等复杂评估问题中应用。

本项成果不仅为决策分析和信息融合领域提供了新工具，也发展了概率语言环境下的测度理论与不确定推理方法。成果完成人完成 1 本英文学术专著，在该领域发表论文 6 篇，其中 ESI 高被引论文 2 篇（被引 812 次）。

**主要创新点 3：针对残缺的定量型审计数据，提出了基于犹豫模糊理论的信息融合方法、分类算法和智能决策算法，既实现了不确定数据的融合，扩大了审计数据取证范围，也创新了适用于非平衡审计数据分析的人工智能技术。**

新时代的审计业务通常需要汇集大量的相关业务部门数据甚至跨行业跨领域数据。由于各部门信息系统的设计缺乏统筹，信息系统记录数据的格式、内容和规范不统一，加上潜在的人为因素，导致数据完整性极差，数据记录不完整。如何从体量庞大的残缺数据中发掘有价值的审计证据是一个难题。项目完成人利用并发展了基于犹豫模糊集的不确定推理技术，提出了适用于多源不确定数据集的信息融合方法和人工智能算法。项目成果不仅极大地扩展了可用于审计分析的数据范围，也为审计数据分析提供了高效的机器学习算法。项目成果收到学术界的高度关注，已在决策支持系统、中文情感分类等领域得到了成功应用。本项创新点的主要成果如下：

(1) 创建了广义犹豫模糊集的基础理论体系。犹豫模糊集允许以模糊多集的形式表述主观或客观不确定信息。面对包含数据残缺等不确定性的多源审计数据集，项目完成人基于直觉模糊理论，扩展了犹豫模糊集原始模型，允许从积极、消极（或乐观、悲观，正面、反面）两个相反的角度表示不确定信息，定义了广义犹豫模糊集的序关系、运算法则和不确定测度，系统构建了广义犹豫模糊集的基础理论体系。该模型自提出以来，受到了学术界的持续关注



注和引用。

(2) 提出了基于广义犹豫模糊集的信息融合理论和决策支持技术。在上述模型的基础上，项目完成人提出了扩展原理，允许将现有的信息融合技术应用于该模型，并构建了决策支持应用的系统原型，为信息融合提供了良好的可扩展性，且支持正、反两个角度进行不确定性推断。该技术为残缺审计数据推断提供了有效途径，其伴随的不确定度量指标方便审计人员直观的理解审计风险。当风险小于给定阈值时，审计人员可以对基于此技术融合的数据发表审计结论。因而，该技术极大地增强了审计数据的可用性。

(3) 创新地提出了基于犹豫模糊推理的机器学习算法。审计数据分析通常可以看作是一个非平衡的二类（合规与不合规）分类问题。项目完成人基于上述犹豫模糊集模型和信息融合理论，分别提出了基于定量数据的非平衡二类分类的集成学习算法和基于语言文本的非平衡二类分类的集成学习算法。相比同期其他主流算法，提出的算法充分考虑数据不确定性和基分类器的不确定性，利用提出的动态加权融合方法，实现了对数据更精确的分类。

本项成果为不确定决策分析提供了新模型，也丰富了基于不确定数据的机器学习算法。成果完成人在该领域发表论文 5 篇，其中 ESI 高被引论文 1 篇，在 Web of Science 中被引 200 次，形成发明专利 5 项，软件著作权 7 项。

## 社会效益

**支撑科技强审。**本项目成果被公安部审计督察局应用采纳，获得“智慧审计”项目立项，对推动全国公安审计工作有重要影响；被江苏省审计厅电子数据处采纳，在组织开展审计业务电子数据的采集、融合和综合分析方面进行了相关应用；南京市审计局、盐城市审计局、连云港审计局、海安市审计局采纳，对审计人员能够熟练运用对数据技术，在审计监督中发挥作用具有十分重要意义。本项目成果被珠海市审计局采纳并应用，签订了关于《智能化审计局平台数据采集分析及其可视化系统的开发》的技术开发合同，集成本项目成果，实现对审计大数据的智能分析；签订了关于《数字化智能审计理论与关键技术研究》的技术开发合同，基于本项目理论成果，合作研究智能审计关键技术在地地方审计局 的国家审计业务中的全面应用。

**助力政务服务。**2019 年 12 月，本项目理论成果被南京市政务服务管理办公室采纳并应用，全面应用于对政务服务数据的采集、清洗、融合与分析，合作完成了《南京市政务服务年度数据报告》。成果对提高南京市政务服务水平，推进简政放权，提高政府治理能力具有重要意义。

**辅助打击金融犯罪。**2018 年 1 月，本项目理论成果被南京市公安局采纳，并签订《南京市公安局与南京审计大学战略合作意向书》。本项目理论成果在南京市公安局的应用有效提升了公安经侦工作的情报分析能力、决策能力以及指挥管理能力，有效推动互联网金融犯罪打击防范工作，对全面推进实施公安大数据战略、实现社会治理现代化发挥了示范引领作用。



## 简介

### 国网智能电网研究院有限公司

#### 江苏省计算机学会常务理事单位

国网智能电网研究院有限公司（以下简称“智研院”），是国家电网有限公司直属科研单位，2013年正式成立，2016年2月更名为全球能源互联网研究院，2017年10月改制为全球能源互联网研究院有限公司，2021年12月更名为国网智能电网研究院有限公司。

国网智研院定位于公司前沿技术创新中心和先进装备研发中心，重点开展新型电力系统基础共性、战略性、前沿性技术研究，新型高端装备、系统、器件、新材料等开发，建设跨领域、跨专业的国际化科研平台，努力打造世界一流的高端技术研发基地、重大成果输出基地和高端人才培养基地。

国网智研院共设置6个职能部门、2个业务支撑部门，7个研究所及欧洲研究院。截至2022年2月底，共有全口径员工800余人，平均年龄36岁，硕士研究生及以上学历占比77.9%，高级职称占比超过35%。拥有中国工程院院士1人，中国科学院院士1人，德国国家工程院院士1人，中央直接联系高级专家1人，享受政府津贴专家4人，“新世纪百千万人才”2人，国家科技创新领军人才2人。2018年获批国家创新人才培养示范基地。

智研院构建了“高性能电工材料、大功率电力电子器件、先进输变电技术装备、新型储能与综合能源系统、智能感知与量测、先进计算与人工智能、信息通信及网络安全”7个主攻方向。拥有先进输电技术国家重点实验室、输配电及节电国家工程研究中心；2个北京市重点实验室；电力系统人工智能等9个公司实验室，重大仪器设备200余台（套）。建设了高压直流输电技术与装备国家创新团队，2019年获国资委中央企业优秀科技创新团队，2021年获全国专业技术人才先进集体，拥有特高压直流电缆等6个公司科技攻关团队。

善学者尽其理，善行者究其难。展望未来，智研院将以习近平新时代中国特色社会主义思想 and 党的十九大精神为指导，积极践行创新驱动发展战略，聚焦能源转型和未来电网发展需求，以加快能源转型、推进“碳达峰、碳中和”为目标，深入实施“新跨越行动计划”，加强科技顶层设计，攻克关键核心技术，落实科研改革措施，持续激发创新活力，努力打造能源领域国家战略科技力量，为建设具有中国特色国际领先的能源互联网企业提供坚强支撑，为全面建设社会主义现代化国家作出新的更大贡献！