

2020年北京市科学技术奖提名公示内容（公告栏）

一、项目名称

开放环境下数字伪造内容检测关键技术与服务平台建设

二、候选单位

1、中国科学院计算技术研究所;2、人民网股份有限公司;3、杭州中科睿鉴科技有限公司

三、候选人

1、曹娟;2、郭俊波;3、高科;4、唐胜;5、李锦涛;6、柳轩;7、乔会朋;8、何雪姣;9、李慧芬;10、吕永标;11、谢添;12、刘浩远

四、项目简介

人工智能伪造技术的迅速发展，使得各种以假乱真的虚假伪造信息泛滥，打破了原有的风控审核技术体系，给国家安全，金融风控，媒体内容风控等多个重要领域带来急迫需要解决的新挑战。因此，数字内容伪造检测被公安部、网信办等安全部门列入重点攻关方向，同时科技部，基金委等科技部门也设立了重点方向对此进行技术攻关，具有重要的学术意义和应用价值。

本项目从系统平台搭建、核心算法攻克，企业落地应用三个层面进行布局，在国内较早且持续地开展多模态虚假伪造信息检测技术和系统研发。针对真实应用场景中面临的信息不完整性，任务不确定性，环境强对抗性问题展开科技攻关，主要技术创新点如下：

（1）基于可信度传播网络的虚假新闻早期检测：在事件发展早期进行虚假检测价值非常高，但可用数据稀疏且不完整，容易导致模型误判。我们提出基于正反观点博弈的可信度传播网络，提高了谣言的早期检测准确率。

（2）基于多尺度视觉模型的虚假伪造图片检测：视觉内容伪造方法各种各样，很难从视觉上对虚假进行明确定义和建模。我们开创性的提出从造假痕迹到造假意图的多尺度检测框架，实现了虚假图像检测精度的大幅度提升。

(3) 基于多模态融合的虚假信息检测：真实场景中发布的虚假信息具有很强的对抗性，多模态融合是最有效的解决方案。我们提出基于注意力机制的多模态融合模型，对文本，视觉和社交特征进行动态融合，在线上系统中达到了最优性能。

(4) 多媒体虚假伪造检测平台构建：虚假伪造数据属于小样本异常数据，需要长期积累。因此本项目从 2014 年开始搭建了一个集成采集、算法研判、标注、对外服务全流程的系统级平台，24 小时不间断运行至今 6 年多，不仅形成了行业内的数据壁垒，同时也为后续产品化奠定了基础。

基于上述创新成果，发表论文 50 余篇，谷歌引用 1750 次。申请专利 32 项，10 项已授权，在虚假信息检测方向形成较高知识产权壁垒。研发了“AI 识谣”虚假新闻检测平台，“数字内容伪造检测风控系统”和“疫情可疑新闻实时发现系统”三个产品。已在公共安全、金融风控、媒体内容风控和疫情防控四个领域进行产业化落地。尤其在北京地区，通过为网信办，广电总局，公安部，新华社等国家部委提供网络安全审核技术支撑服务，发挥了北京在国家安全领域的战略地位；通过与人民网、新华网、华平智保等企业合作进行媒体内容风控审核和金融风控审核，产生了较大的经济效益。

五、经济效益

5.1 直接经济效益

直接经济效益汇总（金额单位：万元）

年 份	项目收入	项目利润	上缴的税收	节支总额
2019 年	703			
2018 年	266.22			
2017 年	127.8			
累 计	1097.02			
效益产生单位				
第*候选单位	单位名称			
1	中国科学院计算技术研究所			

五、经济效益

候选单位经济效益（金额单位：万元）

候选单位排序	1	单位名称	中国科学院计算技术研究所		
年 份	项目收入	项目利润	上缴的税收	节支总额	
2019 年	703				
2018 年	266.22				
2017 年	127.8				
累 计	1097.02				
<p>各栏目的计算依据（限 800 字）</p> <p>2017 年总收入 127.8 万元，主要是在中国电子科学研究院、国家广电总局等机构的技术开发类项目产生的收入</p> <p>2018 年总收入 266.22 万元，主要是在国家计算机网络与信息安全管理中心、新华通讯社等机构的技术开发类项目产生的收入。</p> <p>2019 年总收入 703 万元，主要是在国家计算机网络与信息安全管理中心、北京寺库有限公司等机构的技术开发类项目产生的收入。</p>					
<p>声明：我单位确认以上财务数据真实可靠，同意全力配合后期经济效益数据抽查工作，并愿意承担因此产生的相关责任。</p> <p style="text-align: right;">效益产生单位财务专用章</p> <p style="text-align: right;">年 月 日</p>					

五、经济效益

5.2 推广应用情况

本项目以多媒体虚假伪造信息检测技术为核心，形成了“AI 识谣”虚假新闻检测平台，“疫情可疑新闻实时发现系统”，和“数字内容伪造检测系统”三个产品。并在公共安全、金融风控、媒体内容风控和新冠疫情四个领域进行产业化落地，开始进行大规模商业推广。具体推广情况如下：

(1) 安全领域：2019 年起在国家互联网应急中心和南京市网信办等监管机构上线虚假伪造信息检测系统，辅助监管部门进行争议性信息、虚假信息的筛查和提前感知，并制定分级分领域的有效治理措施。迄今为止，日均检测的数据量超过 20 万条，已预警 7 万条高可疑线索；

支撑材料【附件 17：国家互联网应急中心项目合同】

(2) 媒体领域：2016 年起分别在新华社、新华网和人民网部署新闻素材真伪审核 SaaS 服务，辅助采编人员提高新闻审核和新闻生产效率。本项目和人民网合作完成了“人民网-中科睿鉴新闻可信度识别支撑系统”，将本项目的核心技术嵌入到人民网新闻监测和内容管理系统中，目前融入了人民网大部分核心频道的新闻内容生产流程，极大地提高了新闻的准确性以及内容生产的效率。

支撑材料【附件 27-34：人民网应用证明，新华社应用证明，新华网应用证明，人民网项目验收单，人民网 2020 两会期间创新成果展示界面】

(3) 金融领域：2020 年开始分别在北京华平智保和中国银联上线证件材料真伪审核平台，围绕投保、放贷、理赔等审核业务，将图像视频的 PS 伪造检测核心技术嵌入系统中，每天审核单据 2 万条以上，大大提高了审核效率。目前已经进一步向各大银行，保险和医保平台进行大规模推广。

支撑材料【附件 35：北京华平智保应用证明，附件 25：商务合同，项目收入发票】

(4) 疫情服务：研发的“新冠疫情可疑新闻实时检测系统”于 2020 年在南京舆情研究院部署，面向公众提供及时的可疑新闻查询服务；同时为南京联防联控指挥部每天提供实时信息，在南京疫情期间的互联网治理做出了积极贡献。

支撑材料【附件 33：南京舆情研究院应用证明，“新冠疫情可疑新闻实时检测系统”界面截图】

(5) 公众服务：本项目于 2017 年起与腾讯合作，联合开发了“AI 精准识谣”微信小程序，面向公众提供健康、科普领域的可疑新闻查询，同时联合腾讯科普协会的 2000 名专家形成人机协同的谣言线索发现和专家辟谣联动流程。从 2018 年起，通过中央网信办开通“AI 识谣”头条号，为公众提供及时的辟谣信息。至今已经发布 800 多次辟谣信息，累计推荐阅读量 1.3 亿。

支撑材料【附件 37-38：计算所-腾讯“AI 精准识谣”开发合同；“AI 精准识谣”小程序界面截图，附件 40：“AI 识谣”头条号截图】

六、主要知识产权支撑材料目录（限 10 个）

序号	知识产权类别	名称	国（区）别	授权号	授权公告日	发明人	权利人
1	发明专利权	一种基于目标检测的大规模图像检索方法与系统	中国	ZL201710574741.4	2020-06-05	唐胜, 肖俊斌, 李锦涛	中国科学院计算技术研究所
2	发明专利权	一种有争议性新闻线索自动发现的方法及系统	中国	ZL2015104351054	2019-10-18	曹娟, 张勇东, 张俊强, 李锦涛	中国科学院计算技术研究所
3	发明专利权	融合局部信息的场景分割修正方法与系统	中国	ZL201710650541.2	2020-06-22	唐胜, 张蕊, 李锦涛	中国科学院计算技术研究所
4	发明专利权	基于微博内容的关键词挖掘方法及系统	中国	ZL201410768704.3	2018-09-07	高宝强, 吴波, 曹娟, 张勇东, 李锦涛	中国科学院计算技术研究所
5	发明专利权	一种面向用户生成内容的分布式采集方法与系统	中国	ZL2015101029800	2018-01-09	张勇东, 吴波, 曹娟, 郭俊波, 李锦涛	中国科学院计算技术研究所
6	发明专利权	新媒体智能在线编辑方法、装置及网络信息发布平台	中国	ZL2015104971273	2019-07-30	曹娟, 张勇东, 金志威, 李锦涛	中国科学院计算技术研究所
7	发明专利权	新闻线索个性化推送方法及系统	中国	ZL201510550175	2019-07-30	曹娟, 张勇东, 张俊强, 李锦涛	中国科学院计算技术研究所

				4				
8	发明专利权	一种微博信息采集方法及系统		中国	ZL2013102981197	2016-08-10	吴波, 曹娟, 张勇东, 李锦涛	中国科学院计算技术研究所
9	发明专利权	融合全局信息的场景分割修正方法与系统		中国	ZL201710650525.3	2020-09-11	唐胜, 张蕊, 李锦涛	中国科学院计算技术研究所
10	发明专利权	一种生成描述图像内容的自然语言的方法与系统		中国	ZL2017112652103	2020-01-21	唐胜, 李灵慧, 张勇东, 李锦涛	中国科学院计算技术研究所
序号	知识产权类别	名称		标准类别	标准编号	标准发布日期	标准起草单位	标准起草人
序号	知识产权类别	论文(著作)名称	刊名/出版社	年卷期页码	发表时间(年月日)	通讯作者(含共同)	第一作者(含共同)	论文全部作者
1	论文	Exploring the Role of Visual Content in Fake News Detection	Lecture Notes in Social Networks . Springer (Springer 丛书系)		2020-01-01	曹娟	曹娟	曹娟, 亓鹏, 盛强, 杨天韵, 郭俊波, 李锦涛

		n(综述)	列)					
2	论文	Integrating Semantic and Structural Information with Graph Convolutional Network for Controversy Detection	The Association for Computational Linguistics (ACL, CCF A类)		2020-06-01	曹娟	钟雷	钟雷, 曹娟, 盛强, 郭俊波, 王子昂
3	论文	News verification by exploiting conflicting	AAAI 2016 (CCF A类)		2016-02-12	张勇东	金志威	金志威, 曹娟, 张勇东

		social viewpoints in microblogs						
4	论文	Novel Visual and Statistical Image Features for Microblogs News Verification	IEEE Trans. Multimedia (SCI 国际期刊)		2017-08-01	张勇东	金志威	金志威, 曹娟, 张勇东, 周建设, 田奇
5	论文	News Credibility Evaluation on Microblog with a Hierarchical	IEEE International Conference on Data Mining		2015-08-19	张勇东	金志威	金志威, 曹娟, 姜育刚, 张勇东

		Propagat ion Model						
--	--	--------------------------	--	--	--	--	--	--

七、国家法律法规要求的行业批准文件目录（限 10 个）

序号	审批文件名称	产品名称	审批单位	审批时间	批准有效期	申请单位

八、经济效益支撑材料目录（限 10 个）

序号	候选单位	支撑材料种类	名称 (限 20 字)	支付方	效益产生时间	项目收入(万元)	备注
1	杭州中科睿鉴科技有限公司	技术服务合同（协议）	合同 1-华平智保-伪造图像检测合同	北京华平智保科技有限公司	2020-11-03	500	
2	中国科学院计算技术研究所	技术开发合同（协议）	合同 2-新华社-虚假新闻检测合同	新华社	2018-01-22	86.4	
3	中国科学院计算技术研究所	技术开发合同（协议）	合同 3-网安中心-虚假新闻检测合同	国家计算机网络与信息安全管理中心	2019-11-27	78	
4	中国科学院计算技术研究所	技术开发合同（协议）	合同 4-北京寺库-图像检测合同	北京寺库商贸有限公司	2019-08-20	160	
5	中国科学院计算技术研究所	技术咨询合同（协议）	合同 5-中央网信办-人工智能安全分析合同	国家互联网信息办公室	2018-12-19	20	
6	中国科学院计算技术研究所	技术开发合同（协议）	合同 6-网安中心-特定人物图像检测合同	国家计算机网络与信息安全管理中心	2020-09-01	110	
7	中国科学院计算技术研究所	技术开发合同（协议）	合同 7-网安中心-音视频分析合同	国家计算机网络与信息安全管理中心	2020-07-08	40	
8	杭州中科睿鉴科技有限公司	技术服务合同（协议）	合同 8-南京报业-虚假新闻检测合同	南京报业集团	2020-10-29	28	
9	杭州中科睿鉴科技有限公司	技术服务合同（协议）	合同 9-美亚柏科-伪造图像检测合同	厦门美亚柏科信息股份有限公司	2020-10-29	25	
10	中国科学院计算技术研究所	技术开发合同（协议）	合同 10-五十七所-文本内容分析合同	五十七所	2019-05-23	28	

九、提名意见

数字内容伪造检测是近几年人工智能生成技术的发展给网络安全带来的急迫需要解决的新挑战，被公安部、网信办等部委列入重点攻关方向，同时也列入了中科院计算所“十四五规划”重点培育方向。该成果从数据平台搭建、核心算法攻克，企业落地应用三个层面进行布局，在国内较早且持续地开展该领域探索。项目组针对真实应用场景中面临的信息不完整性、任务不确定性、环境强对抗性问题展开科技攻关，开发了“AI 识谣”虚假新闻检测平台，“伪造图片和视频检测系统”，和针对疫情开发了“新冠肺炎可疑新闻实时识别系统”，在公共安全、金融、媒体和疫情防控四个领域大规模应用，尤其在北京地区企业落地成果较多，产生了很大的社会效益和经济效益。

提名该项目为北京市科学技术奖科学技术进步奖（类别：技术开发类）（一等奖和二等奖）