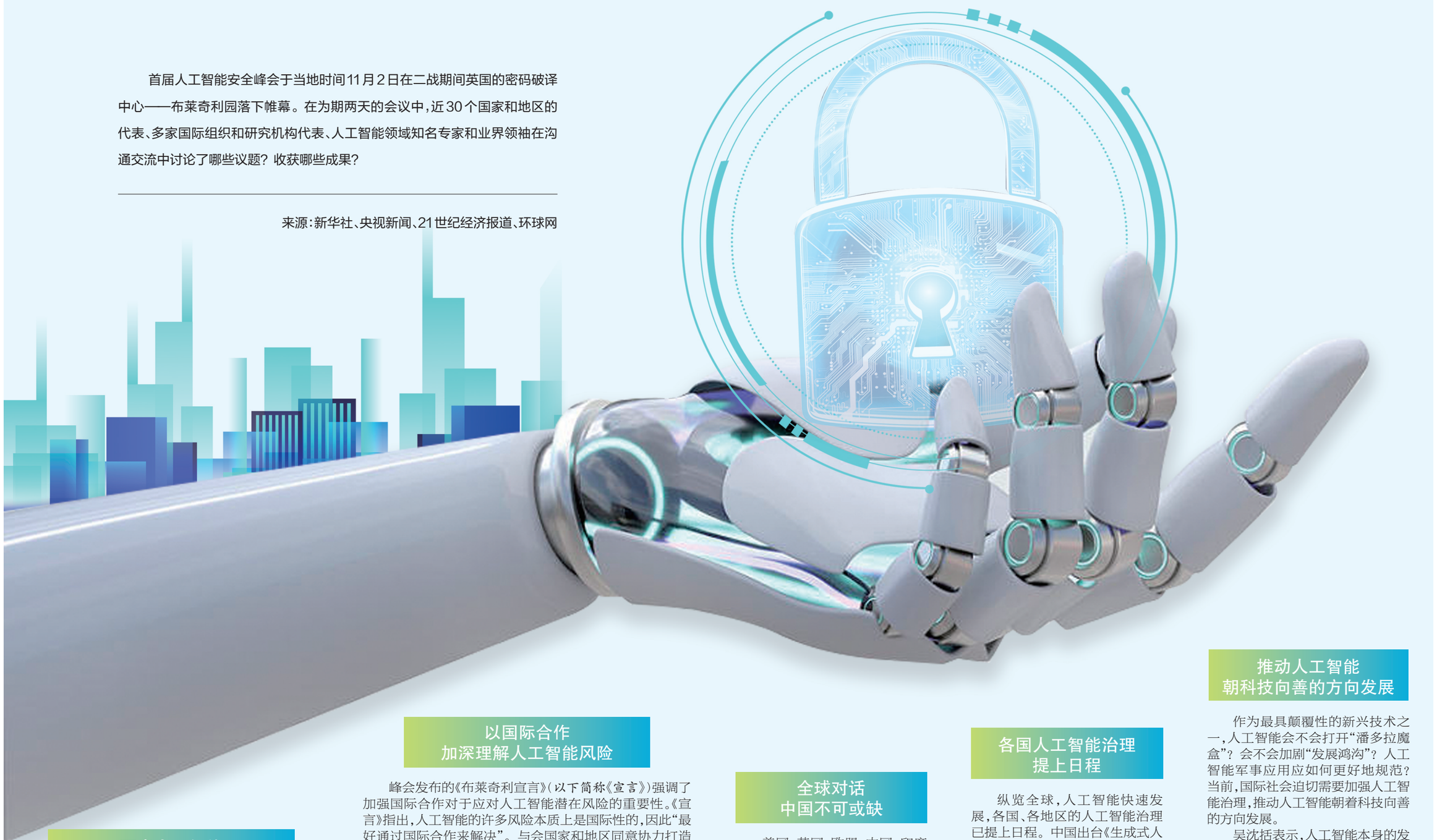
首届人工智能安全峰会开得怎样？

首届人工智能安全峰会于当地时间11月2日在二战期间英国的密码破译中心——布莱奇利园落下帷幕。在为期两天的会议中，近30个国家和地区的代表、多家国际组织和研究机构代表、人工智能领域知名专家和业界领袖在沟通交流中讨论了哪些议题？收获哪些成果？

来源：新华社、央视新闻、21世纪经济报道、环球网



两类人工智能 五大目标

会议期间，美国、英国、欧盟、中国、印度等多方代表就人工智能技术快速发展带来的风险与机遇展开讨论。来自全球的约150名代表重点关注前沿人工智能的潜在风险，同时也考虑狭义人工智能的风险。前沿人工智能是指处于当前能力前沿的通用人工智能模型，可能带来技术滥用、失去控制等风险；狭义人工智能是指能力限于特定领域的人工智能模型，它们也具有潜在风险，例如生物工程领域人工智能模型可能被用于开发生物武器等。

会议主要围绕五大目标展开讨论：第一，对前沿人工智能带来的风险和采取行动的必要性达成共识；第二，推进前沿人工智能安全国际合作，包括如何最好地支持国家框架和国际框架等；第三，推动各相关组织采取适当措施，以加强前沿人工智能安全；第四，寻求人工智能安全研究的潜在合作领域，包括评估模型能力和制定用于支持治理的新标准；第五，探讨如何确保人工智能安全发展，使人工智能在全球范围内被妥善运用。

根据议程安排，各方代表在峰会开幕当天主要就前沿人工智能带来的风险类型、不同参与方在应对这些风险中的作用、人工智能在不同领域面临的重大发展机遇等展开跨学科对话和讨论。在会议第二天，各方代表主要讨论了人工智能的影响、如何有效合作以及如何进一步实现确保全球人工智能安全的使命。

【链接】

业内人士：加快推动人工智能伦理规范落地

人工智能治理并非只在本届峰会提及，在近日举行的博鳌亚洲论坛全球经济发展与安全论坛第二届大会上，业内人士也围绕人工智能治理与可持续发展进行探讨。

“人工智能带来的数据泄露、个人隐私侵犯、虚假信息泛滥等问题不容忽视。”国家开发投资集团特级专家赵建强认为，人工智能治理的紧迫性越来越强。

解决人工智能治理难题需要加强技术创新。目前，多家科技企业正在展开探索，如商汤科技研发“深伪检测+数字水印”技术，通过技术手段防止内容被篡改或编辑，从源头上防范AI生成虚假内容，打造安全可信的人工智能。

上海交通大学中国法与社会研究院院长、亚洲科技促进可持续发展目标联盟副主席季卫东认为，人工智能设计过程中应嵌入一种风险防范机制，通过其他的人工智能系统来对人工智能进行制衡。

“人工智能治理过程中，需要形成具有广泛共识的标准规范。”季卫东认为，应尽快确立具体的技术标准与监管规范，尤其要加快推动人工智能伦理规范落地。

中关村知识产权战略研究院执行院长、己任律师事务所合伙人何菁以应对人工智能带来的知识产权纠纷举例称，技术界、法律界和内容生产者三方需要合力保护创作者的利益，防止知识产权纠纷成为人工智能发展的“拦路虎”。

与人工智能相关联的大数据跨境流动带来的数据安全问题正不断凸显。中关村实验室首席科学家云晓春认为，在推进数据开放共享的同时，需要同步探索治理与监管体系，形成公平合理的数据流动规则，共同应对数据流动带来的风险和挑战。赵建强建议，要探索基于数据、算法、技术共享的合作机制，制定推动人工智能可持续发展的国际性标准和框架。

以国际合作 加深理解人工智能风险

峰会发布的《布莱奇利宣言》（以下简称《宣言》）强调了加强国际合作对于应对人工智能潜在风险的重要性。《宣言》指出，人工智能的许多风险本质上是国际性的，因此“最好通过国际合作来解决”。与会国家和地区同意协力打造一个“具有国际包容性”的前沿人工智能安全科学研究网络，以对尚未被完全了解的人工智能风险和能力加深理解。

英国政府发表声明称，这是世界首个多个国家和地区达成的人工智能安全领域宣言，参会代表就前沿人工智能技术发展面临的机遇、风险和采取国际行动的必要性取得共识。

本次峰会及其成果引发媒体、学界、业界专家的高度关注和积极评价。“《宣言》是在管理风险和实现前沿人工智能系统的好处方面迈出的重要一步。”英国剑桥大学“人工智能：未来和责任”项目负责人肖恩·奥黑盖尔塔格表示，《宣言》强调了开发前沿人工智能的参与者所肩负的“特别重大的责任”，包括拥有充足资源的科技公司以及他们所在国家的政府等。

值得注意的是，作为本次峰会的重要成果之一，《宣言》由来自中国、印度、美国和欧盟等28个国家和地区的代表签署通过。签署方一致认为，迫切需要通过新的全球联合努力来了解和集体管理潜在风险，以确保人工智能安全。以安全、负责任的方式开发和部署人工智能，造福全球社会。

但也有观点认为，本次峰会虽然达成了指导国际合作的原则性宣言，但并未商定国际合作路线图，也没有就各国如何在应对人工智能风险问题上开展合作提出具体措施，未来需要采取更积极主动的国际合作行动。“《宣言》提出AI全球治理问题，呼吁各国关注，努力做出更进一步的治理探索，这是目前能够看到的一个价值。”中国互联网协会研究中心副主任吴沈括说。

在他看来，不宜把《宣言》意义看得过高。“《宣言》反映出来的对于人工智能治理的认知处于相对初期的阶段，更多是在说需要治理、呼吁共同合作识别风险、构思解决方案。相较于中国发布的《全球人工智能治理倡议》，两个文件在发展阶段、认知阶段存有代差，《宣言》并没有对人工智能治理的核心关切和风险进行直接提炼，也未给出解决思路。”

全球对话 中国不可或缺

美国、英国、欧盟、中国、印度等多方代表在两天会期内，就人工智能技术快速发展带来的风险与机遇展开讨论。英国图灵研究所、中国科学院、经合组织等众多机构，深层思维、谷歌、腾讯、阿里巴巴等企业，以及马斯克、杨立昆、约舒亚·本乔、杰弗里·欣顿等行业专家也出席了峰会。

本届峰会上，中国代表团的出席引人关注。多位专家强调，中国作为在人工智能研发领域领先的国家之一，在应对人工智能风险和机遇的全球讨论中不可或缺。

世界知识产权组织数据显示，仅2022年一年，中国机构人工智能专利申请数量就多达29853项，占当年全球人工智能专利申请总量的40%以上。

美国外交政策智库卡内基国际和平研究院院长蒂诺·奎利亚尔对媒体表示，只有中国参会，才能证明这是一次真正的全球对话。

美国硅谷企业家埃隆·马斯克在峰会闭幕后表示：“如果中国不是（峰会）参与者，那就毫无意义。”他认为，美国、英国和中国等各方在人工智能安全方面取得共识有利于应对相关问题。

中国科技部副部长吴朝晖率团参加本届峰会，并在开幕式全体会议上发言。与会期间，中方代表团参与人工智能安全等问题讨论，积极宣介中方提出的《全球人工智能治理倡议》，并与相关国家开展双边会谈。

中方表示，愿与各方一道就人工智能安全治理加强沟通交流，为推动形成普遍参与的国际机制和具有广泛共识的治理框架积极贡献智慧，切实落实全球发展倡议、全球安全倡议和全球文明倡议，促进人工智能技术更好造福人类，共同构建人类命运共同体。

对此，清华大学新闻学院元宇宙文化实验室主任沈阳表示，《全球人工智能治理倡议》在应对AI带来的挑战、全球协作与治理、发展中国家的代表性和发言权、推动国际合作与交流等方面具有重大意义。倡议认识到，强调AI治理攸关全人类命运，是世界各国面临的共同课题。倡议强调了全球协作和共同治理的重要性，倡导各国在AI发展和治理方面享有平等的权利和机会。这是在全球化背景下，寻求解决全球性问题的重要一步。

各国人工智能治理 提上日程

纵览全球，人工智能快速发展，各国、各地区的人工智能治理已提上日程。中国出台《生成式人工智能服务管理暂行办法》为生成式AI立规并鼓励发展；近期发布《全球人工智能治理倡议》从发展、安全和治理三个板块出发，为全球人工智能治理提出了11项主张。欧盟的《人工智能法案》则在拉锯多个回合之后，进入冲刺阶段。

美国总统拜登则于近期签署了《关于安全、可靠和可信地开发和使用人工智能的行政命令》，并指示商务部设立美国人工智能安全研究所引领人工智能安全工作。此次峰会上，美国商务部部长吉娜·雷蒙多表示，人工智能安全研究所将评估前沿AI模型已知和新出现的风险。

“关于人工智能治理，从全球来看，有两个大的方向，一是强调安全的治理思路，包括加强政府监管实现安全和通过行业规则实现安全；二是强调人工智能的发展，通过市场驱动发展或通过国家规划驱动发展。”吴沈括说。

AI治理全球竞速的背后，或许离不开对AI治理话语权这一问题的回应。

“各国争夺AI领域的话语权其实基于多个要素的综合作用。”对外经济贸易大学数字经济与法律创新研究中心执行主任张欣分析，首先，人工智能愈发成为各国经济转型的元引擎，具有重要的驱动作用。因此，掌握人工智能治理领域的话语权意味着可以更好地为本国的企业和产业创造有利的运行环境。其次，人工智能的竞争早已从单纯的技术和应用竞争扩展到治理规则的竞争。

“可以说，在这个领域，出现技术赛道、产品赛道和治理赛道多轨并行的局面。因此，在全球人工智能治理框架迅速形成的过程中，积极发挥影响力具有重要的战略意义。”张欣说。

对于中国而言，在AI全球治理过程之中，又该扮演何种角色？

吴沈括认为：“时代已经发生变化，当前并不是说有一个人工智能国际规则需要中国参与融入，现实状态是，中国本身就是国际规则的推动者和重要的共同制定者。人工智能时代，国家地位和力量对比发生了根本性的变化，我们将会是规则共同的缔造者。”

推动人工智能 朝科技向善的方向发展

作为最具颠覆性的新兴技术之一，人工智能会不会打开“潘多拉魔盒”？会不会加剧“发展鸿沟”？人工智能军事应用应如何更好地规范？当前，国际社会迫切需要加强人工智能治理，推动人工智能朝着科技向善的方向发展。

吴沈括表示，人工智能本身的发展不仅是一个技术问题，它还会深度改变我们的生活、生产方式以及未来发展的图景，包括我们的技术基础，经济发展的模式，乃至国家治理和国际关系。但需要注意的是，一方面部分国家和地区面对人工智能的快速发展，存在认知、能力以及治理的不足；另一方面部分国家利用技术优势推行技术霸凌、技术霸权，这种行径严重损害了其他国家和人民的发展利益。

中方正是在此背景下，提出《全球人工智能治理倡议》，围绕人工智能的发展、安全和治理三方面系统清晰阐述了中国路径和中国方案。倡议强调面向他国提供人工智能产品和服务时，应尊重他国主权，严格遵守他国法律，接受他国法律管辖；发展人工智能应坚持“智能向善”的宗旨，各国尤其是大国对在军事领域研发和使用人工智能技术应该采取慎重负责的态度；发展人工智能应坚持相互尊重、平等互利的原则，反对利用技术垄断和单边强制措施制造发展壁垒，恶意阻断全球人工智能供应链等。

外交部军控司司长孙晓波指出，《全球人工智能治理倡议》坚持发展与安全并重的系统思维，反对以意识形态划线或构建排他性集团，恶意阻挠他国人工智能发展，引导人工智能朝着有利于人类文明进步的方向发展。强调安全可控、隐私保护、公平和非歧视，集中反映了各方对人工智能安全的主要关切，也为相关国际讨论和规则制定提供了蓝本，反映出我们在人工智能领域成熟的治理经验。

此外，《全球人工智能治理倡议》特别呼吁要增加发展中国家的代表性和发言权，开展面向发展中国家的国际合作与援助。这其实是对目前“智能鸿沟”不断扩大的一个回应，跟一些国家主张的“小院高墙”做法形成了鲜明对比。努力弥合“智能鸿沟”，并确保“智能红利”惠及各国，这是中方在人工智能治理问题上的立场，也是国际社会在制定相应标准规范时不容忽视的重要考量。

日前，联合国宣布成立“人工智能高级别咨询机构”，两名中国学者入选机构成员。对此，孙晓波表示，作为负责任的人工智能大国，中方支持在联合国框架下讨论人工智能治理，推动形成具有广泛共识的治理框架和标准规范。