2024世界人工智能大会暨人工智能全球治理高级别会议邀请各国政府和产学研代表出席，打造AI全球治理的高级别“议事厅”。

AI技术和产业革命风起云涌，但治理短板值得全球共同关注——

立即行动，共商共享推进AI向善

■本报记者 **张懿**

每年的世界人工智能大会，都有专门的板块讨论人工智能（AI）治理——也就是以规划、监管、调控等方式驱动AI向善。但像今年这样，以如此高的规格、频度和参与度，全面聚焦这个话题，还是第一次。

昨天在参与今年世界人工智能大会首日活动后，记者意识到，今年这一安排非常明智且重要。可以用在会场里听到的两组数据来解释。

一、2023年，AI发展领先国家所遭遇的AI安全事件数量，是2022年的10倍；二、目前所有算力消耗之中，用来提升AI性能的占99%，用于安全的不到1%。

应该说，AI治理已成为了一个迫在眉睫的问题，亟须引起各方高度关注。

另外，AI治理还有更多非常复杂的因素牵涉其中，包括全球发展不平衡的问题，AI话语权被产业界过多掌控的问题，AI风险应对存在“小圈子”化的问题……这些都凸显了通过共商共享推进AI向善的意义，也是今年世界人工智能大会的主题所在。

治理短板

无论如何，必须承认，AI的技术和产业革命风起云涌，但随之暴露的治理短板已成为值得全球共同关注的问题。联合国秘书长技术特使阿曼迪普·吉尔在大会开幕致辞中说，AI非常重要，但风险又非常高：我们时间窗口有限，必须抓住当下的机会。

上海人工智能实验室主任周伯文提到，大约一年前，全球有数百名AI科学家共同签署一份文件，呼吁要将抵御AI风险作为与疾病大流行、核战争等一样，列为全球优先议题。

应邀在大会上作视频致辞的硅谷企业家加里·马库斯说，应对AI风险就应该像应对气候变化一样，必须立即采取行动。

造成AI风险的重要原因，是技术本身的局限，特别是大模型，由于它建立在统计学基础上而非严格的因果推理，因此，其生成的内容，有时会隐含着幻觉、偏见、歧视等内容。正如马库斯所说，如果AI只是杜撰了一头名为“Jumbo”的大象在1959年完成了横渡英吉利海峡壮举的天方夜谭，那可能只是傻得可爱；但实际上，假如这种情况发生在国际政治军事领域，那就会在短期内制造出巨大危机。

中国科学院自动化研究所人工智能伦理与治理中心主任曾毅说，有研究显示，目前最强大的AI模型，在回答伦理道德问卷时，与人类一致的只有60%，换句话说，接近一半的“认知”与人类是相悖的。

虽然不少人强调要在AI训练时让它更多地理解伦理道德，但曾毅强调，实际上，今天的AI并不存在“理解”这回事，它只是在“处理”。实际上，追溯一个人如何形成道德和伦理感知的过程就会发现，他必须有“自我”的体验，进而具备反思、共情、推己及人的能力，最终形成道德知觉和推理。反观AI，虽然它在和人对话时口口声声“我以为”，但实际上它是没有自我的。而要彻底改变这个局面，就需要从技术底层入手，重塑AI的基础。

错综复杂

不过，在技术问题之外，还有其他一些因素，为管控AI风险带来不必要的难度。

作为联合国高级别人工智能顾问机构39位全球专家之一，曾毅表示，2023年AI安全事件10倍增长背后，还有一个令人吃惊的事实：不同国家反复出现的AI风险，常常是相似的，但大家却总是一错再错：“目前，AI治理在全球形成一些‘小圈子’，给治理造成不利影响。”

曾毅表示，不同国家在AI发展水平上的差异，使得他们的立场和关注点并不一致。发达国家更倾向于向发展中国家推广技术，但在此过程中，他们似乎并不关心对方真正需要的是什么。换句话说，AI治理更应该追求公平公正的格局，而不仅仅是出于利益导向。作为一名AI技术专家，曾毅坦言：“我们真的需要让所有事情都交给AI解决吗？我对此表示怀疑。”

谈到利益对AI治理的影响，马库斯有更深入的分析。虽然本人就来自硅谷，但他毫不掩饰自己的批评态度。他认为，目前美国大型科技公司在AI领域的话语权太大了，看起来他们能够左右一切。而在监管无法充分发挥作用的情况下，是否应该允许AI加速奔跑呢？此外，他还提到了一个硅谷固有的“顽疾”——炒作。马库斯说，硅谷总是在试图让你相信，AGI（强人工智能）已近在咫尺，比如在2030年甚至2027年就会出现。而一旦你真的这么认为了，那么AI眼前所存在的这些问题，就会变得不再重要。而在现实中，有研究显示，大模型可能已经达到了在当前范式下的某种“收益递减点”，也就是说，它的增长速度已经放缓了，比如从GPT-4到Turbo版，似乎并没有太明显的进步。

负起责任

面对AI发展与安全之间的失衡，构建一个多元参与的治理格局就显得尤为重要。昨天的大会上，包括新开发银行行长罗塞夫、联合国工发组织总干事穆勒、联合国秘书长技术特使吉尔在内，有不少嘉宾对中国倡导的AI治理倡议给予非常高的评价，他们都认为，中国提出在联合国框架内构建AI国际治理体系，关注发展中国家的权利和机会，并通过强化AI能力建设，确保他们不掉队的倡议，有助于创造一个AI服务全人类的未来。

也正是在这样一种愿景之下，昨天在大会开幕式上发布的《人工智能全球治理上海宣言》得到了广泛的认同。宣言开宗明义地提出：“我们相信，只有在全球范围内的合作与努力下，我们才能充分发挥人工智能的潜力，为人类带来更大的福祉。”

清华大学人工智能国际治理研究院院长薛澜表示，当今时代，人工智能给一个国家带来的风险，实际上就是全球风险，因此必须得共同努力才能解决。曾毅说，中国作为AI发展领先国家，通过分享自己AI治理的最佳实践，帮助其他国家快速提升，这是负责任大国应有的态度。

除了国际合作，很多人也强调了AI治理多元参与的重要性。图灵奖得主、中国科学院院士姚期智说，AI几乎涉及每一个行业，相应地，会产生各种类型的风险，因此在治理中需要科学家、政府官员、法律界专家、经济学家等的广泛参与。曾毅也建议要在AI治理中形成一个“机制复合体”，政产学研，包括媒体和公众，都应该承担相应的义务。在他看来，如果说前沿研究者关注如何造就更强的AI，但对于AI的风险，公众和媒体的敏锐度可能更高。

马库斯则认为，可以借鉴航空监管体系，构建一个由独立的第三方专业机构，在AI大规模部署前进行测试、审计工作。这样一种机制有效地确保了商业航空的安全性，即便发生事故，也能够快速找出问题。

据记者观察，本次世界人工智能大会期间，与安全治理相关的高层级论坛会议不下10场。未来两天，这个话题的热度预计仍然会像上海的天气一样，陡然上升并保持在高位。无论如何，正如曾毅所说，面对AI风险的挑战，最重要的不仅是达成共识，更要推动实践。他认为，当前AI行业具备贯彻相关倡导和规范的条件，全社会应该敦促业界抓住有限的时间窗口，将这些伦理要求落地。

双轮并行，“发展治理”成核心共识

■本报记者 **史博臻**

2024世界人工智能大会主题为“以共商促共享，以善治促善智”，而在昨天开幕当天举行的“人工智能前沿技术的治理挑战与应对措施”论坛上，专家学者围绕人工智能（AI）前沿技术的发展与治理，探讨治理挑战与应对措施，探索构建人工智能治理的全球框架。

当前，随着通用人工智能的广泛应用和深入研究，前沿人工智能的发展潜力和安全风险更为直观地展现了出来。而“发展治理”已成为核心共识，世界各国需要通过加强对话与互动，以更务实的国际合作，防范人工智能快速发展可能引发的极端风险。

加强法律法规的协调和统一

在人工智能领域，数据安全受到了广泛关注。数据是人工智能的基础，也是数字经济时代的核心资源，随之而来的就是数据安全和隐私保护的重大挑战。

“在全球关注数据安全治理的背景下，我们需要加强国际间的合作与交流，共同探讨数据安全治理的最佳实践，分享经验和教训，加强法律法规的协调和统一。”根据香港科技大学首席副校长、香港生成式人工智能研发中心主任郭毅可的观察，近年来，各国密集出台了涉及人工智能的法律和政策，“目前，全球大约有33个国家和地区已经制定了人工智能相关法律，还有许多国家正在起草、谈判和通过相关法规，彰显了国际社会对数据安全治理的重视。”

把不同的国际数据安全治理框架展开对比可以发现，一些国家采取了专门的法律和法规，来覆盖人工智能发展应用中的广泛问题，而另一些国家则采取了针对不同应用或类型的人工智能，制定了各种规章制度的策略。比如，《欧盟人工智能法案》对高风险的人工智能系统进行了禁止，并对各类风险系统设定了更严格合规的标准。与之相对的是，英国、美国采取了市场驱动的方法，避免过度限制技术进步，注重推动创新。

与此同时，数据安全治理与隐私保护息息相关。郭毅可表示，在人工智能的发展中，大量个人数据被用于训练和应用，为了解决隐私和数据保护问题，各国也纷纷制定了相关规定。比如，英国《人工智能法规白皮书》，要求AI开发者遵守安全、透明、公平、负责任和可解释的原则。

除了数据安全治理制度规则，数据安全领域的国际前沿技术同样备受关注。“比如，区块链技术是另一种有前景的解决方案，可增强数据安全。”他认为，其透明和不可更改的技术特性，能确保数据的完整性、降低数据的授权和篡改风险，因此，基于区块链的解决方案在金融、医疗和供应链管理等行业具有特别价值。

为全球共同治理提供中国实践

以ChatGPT和Sora发布为节点，人类在短短几年间已连续两次感受到了“大模型+生成式AI”的冲击。在清华大学人工智能国际治理研究院院长薛澜看来，总体上，以生成式AI为代表的本轮人工智能浪潮，总体处于“工具化末期、产品化初期”阶段。

梳理人工智能全球治理的进展，“发展治理”成为核心共识。多个国家和地区发布国家基础设施及人工智能投资计划，聚焦可持续算力、高质量数据集等大模型研发的核心生产要素补齐，同时围绕人才、资本、监管等制度要素进行合理配置，逐步形成了“发展型治理”体系。

与机遇并存，前沿人工智能主要面临技术挑战和国际治理挑战。前者包括功能障碍和失控风险等问题；后者主要有治理机制变革与技术发展缺乏同步性，发达国家和发展中国家技术鸿沟难以解决，机制复合体治理系统复杂，共识尚且处于初步阶段等情况。

如何加速构建全球人工智能共同治理体系？薛澜表示，他和团队把研究成果汇编成《全球人工智能治理与中国方案》，从3个方面阐述了相关理念。

首先，求同存异，持续沉淀基于共识的全球人工智能治理理念，不仅要对技术应用保持耐心，更多由市场力量推动；同时要防止过度超前的监管措施压制行业发展预期。其次，承继创新，构建开放包容的全球人工智能治理体系。最后，循序渐进，分步打造综合系统的全球人工智能治理框架。做好4个统筹，即统筹发展与安全，统筹伦理、立法、标准、测评等核心治理板块，统筹国内与国际治理，统筹既有治理经验和治理资源与人工智能发展面临的实际问题。

“中国实践可以为全球人工智能治理注入动力。”薛澜还特别提到，中国互联网行业为人工智能提供了丰富的数据、算力资源和多样的应用场景，同时，中国人工智能核心产业发展迅速，顶尖模型不断取得突破。人工智能加快赋能千行百业，创造大量智能向善的“中国案例”。

今年大会有9位图灵奖、菲尔兹奖、诺贝尔奖得主和88位国内外院士出席，与上千位全球科技、产业界领军人物共同推动国际人工智能领域交流与合作。 均本报记者 **袁婧**摄