Source: *Lianhe Zaobao,* p20
Date: 20 January 2026

当下热点

印尼封禁Grok的启示

白士泮　石建政

印尼封禁Grok的意义在于，它把一个被长期忽视的事实推到台前：生成式AI在相对开放的创意科技生态里得以快速成长，已经不是普通互联网应用，而是一种“生产能力”的扩散，全球监管机构是否应该开始在它的可生成能力方面进行管控。不能只谈技术跃迁，更要把人的尊严写进产品默认值与制度细则。

印度尼西亚近日宣布，马斯克旗下的xAI人工智能应用Grok涉嫌生成色情内容，因此暂时封锁用户使用。随后，马来西亚监管机构也公开表达对类似风险的担忧，并采取限制措施。这一举动引发两种典型反应：支持者认为“必须先刹车再上路”，反对者担心“一刀切”会扼杀创新。若只停留在“封与不封”的立场之争，反而会错过更关键的信号——生成式AI正在把数码治理从事后的“内容治理”推向“事前的能力治理”，东南亚很可能是这一转型最早发生，也最具代表性的试验场。随着各地批判声浪的加大，xAI在上周三表明，已对Grok AI聊天机器人的所有用户实施了限制，禁止进行图像编辑。

首先须要正视的是：这并非单纯的色情内容问题，而是在数码世界里如何保护人权的问题。传统互联网的伤害多依赖传播渠道：一张不雅照、一段偷拍视频，往往源于现实中的侵害，再由平台扩散。生成式AI改变这一链条，它把“制造伤害”的门槛大幅降低，并使伤害可以在极短时间内规模化复制。非自愿的深伪色情内容直接侵害个体的尊严、同意权与社会身份；受害者往往在证据链、取证成本、跨境追责上处于劣势。监管者对这类风险采取强硬动作，并非道德保守主义的回潮，而是公共安全与人权保护在数码时代的必然反应。

监管开始要求产品上线前承担风控义务

但更值得讨论的，是治理逻辑的优化。过去平台治理以“事后删除”为主：发现违规内容，用户举报，平台下架，必要时配合执法。这套机制在社交媒体时代已经疲于奔命，到了生成式AI时代更显捉襟见肘，因为问题不再是“有人发布了什么”，而是“默认系统能生成什么”。当一个工具具备轻易生成高风险内容的能力时，单靠用户举报与事后删除，等同于让社会为技术的默认设置买单。马来西亚监管方对“过度依赖用户举报、护栏不足”的批评恰恰反映这种转向：监管开始要求AI企业在产品上线之前，就应该承担风险控制的义务。

这就是“能力治理”的核心：从管控内容转向制定AI产品能力的边界。治理对象不再只是某一条违规文本或图片，而是模型的可生成空间、风险分级机制、默认设置，以及应急响应体系。AI企业若仍以“我们只是工具”和“用户如何使用不归我们管”的叙事自我辩护，已难以获得社会的信任。因为在生成式AI体系中，可生成空间的产品设计本身就是权力分配：默认开放什么、默认禁止什么、触发什么拦截、是否保留可追溯日志、是否提供受害者快速救济通道，这些都决定风险落到谁身上，也是保护用户的根本。

进一步看，印尼的举动之所以具有象征意义，还在于它揭示东南亚的治理现实：移动互联网普及率高、跨境传播速度快、语言与平台生态多元、监管资源分布不均。生成式AI的风险在这里更容易形成外溢和放大效应。一项AI工具在某国上线，内容却可能在区域内快速传播，受害者与加害者、平台与模型提供者往往分处不同法域。印尼先封、马来西亚跟进，与其说是个别国家的过度反应，不如说是区域性风险的压力测试。

封禁是否就是最优解？未必。封禁更像紧急制动，它能迅速降低风险暴露，但也可能带来两类副作用：一是把用户推向更隐蔽的渠道或替代工具，形成“地下化”活动；二是让监管陷入“永远追着新工具跑”的被动。更可持续的路径，是把理性的“能力治理”具体化为可执行的制度与产品要求，并形成区域协同的共识。

对新加坡及区域决策者而言，这一事件也提供现实启示：创新与安全不应被塑造成零和关系，关键在于“边界”是否可解释、可执行、可追责。东南亚若要避免在全球AI竞赛中被动挨打，就须要在“最低共同标准”上更快形成合力，例如对“非自愿性深伪”的统一定义、跨境通报与下架机制，以及取证协作框架。没有共同标准，监管会碎片化，反而会抬高企业合规成本，最终损害创新生态。

印尼封禁Grok的意义在于，它把一个被长期忽视的事实推到台前：生成式AI在相对开放的创意科技生态里得以快速成长，它已经不是普通互联网应用，而是一种“生产能力”的扩散，全球监管机构是否应该在这个时候，就像核能安全设置，开始在生成式AI的可生成能力方面进行管控。不能只谈技术跃迁，更要把人的尊严写进产品默认值与制度细则。技术越强，越须要把责任前置。真正决定一个社会能否拥抱AI的，不是它能生成多少内容，而是它能否在生成之前就守住底线。

作者白士泮是南洋大学校友学术会顾问与新加坡社科大学客座教授
石建政是新加坡社科大学客座讲师与新加坡国立大学访问学者