UC3M一项研究分析了由人工智能生成的深度伪造品（Deepfake）的特征

大多数由人工智能（AI）生成并通过社交网络传播的深度伪造品（带有虚假超现实再现的视频）是由政治家和艺术家主导的，并且通常与当前新闻周期相关。 这是马德里卡洛斯三世大学 （UC3M）一项研究的结论之一。该研究分析了在西班牙因非法使用人工智能工具而产生的病毒式虚假信息的形式和内容特征。 这一进展表示我们在理解和减轻谣言在社会中产生的威胁方面迈出了一大步。

最近发表于期刊OberCom 上的这项研究中提到，团队通过西班牙事实核查组织（如 EFE Verifica、Maldita、Newtral 和 Verifica RTVE）的验证，对上述提到的虚假内容进行了研究。研究人员，UC3M传播系欧洲传播学博士在读研究生拉盖尔·鲁伊斯·因赛迪斯（Raquel Ruiz Incertis）解释：“我们的目标是确定这些病毒式深度伪造的一系列常见模式和特征，提供一些识别关键，并提出媒体素养建议，以便公民能够甄别错误信息。”

研究人员制定了深度伪造的分类法，这对其识别和消除变得更加容易。 根据研究结果，一些政治领导人（如特朗普或马克龙）是制造吸毒或道德不端活动等虚假内容的主导者。此外，还有相当一部分色情性质的深度造假损害了女性的名誉，尤其是对著名歌手和女演员谣传。 研究人员指出，这些视频通常是从非官方账户分享的，并且通过即时消息服务迅速传播。

深度伪造的激增以及频繁使用人工智能工具操纵的图像、视频或音频是当前炙手可热的话题。鲁伊斯·因赛迪斯表示：“这种事先炮制的骗局在选举前的敏感时期或我们目前正在经历的乌克兰战争以及加沙冲突时期特别具有破坏性。这就是我们所谓的‘混合战争’：战争不仅在实体领域进行，而且还在数字层面进行，而且虚假信息比以往任何时候都更加深入人心。”

无论是国家安全，还是选举活动的完整性，该研究的应用都非常广泛。结果显示：在社交媒体平台上积极应用人工智能可能会彻底改变我们在数字时代保持信息真实性的方式。

研究还强调了提高媒体素养的必要性，并提出了提高公众辨别真实和被操纵内容能力的教育策略。因赛迪斯表示：“许多深度伪造作品可以通过谷歌或必应等搜索器中的图像逆向搜索来识别，甚至有一些工具可以让公民只需点击几下鼠标，即可在传播具有可疑来源的内容之前验证其真实性。 关键是指导他们如何操作。”此外

，她还提供了其他检测深度伪造的技巧，例如注意元素边缘的清晰度以及图像背景的清晰度：如果视频中的动作减慢或存在任何类型的面部表情变形、身体比例失调以及奇怪的光影效果，那么这一切迹象都表明这很有可能是人工智能生成的内容。

此外，研究人员认为亟需立法强制平台、应用程序和软件（如 Midjourney 或 Dall-e）建立一个“水印”标识，以便用户一目了然的知道该图像或视频是否已经完全由人工智能修改或创建。

研究团队采用多学科方法，结合数据科学和定性分析，来检验事实核查组织如何在运作过程中应用人工智能。研究的主要方法是对从上述事实核查机构网站上获取的约三十份出版物进行内容分析，以揭露这些经过操纵或由人工智能制作的内容。

参考书目：

作者：Garriga, M.、Ruiz-Incertis, R. 和 Magallón-Rosa, R.（2024）

《关于深度造假的人工智能、虚假信息以及媒体素养建议》

Artificial intelligence, disinformation and media literacy proposals around deepfakes.

期刊Observatorio（观察者）（OBS*），18(5)

https://doi.org/10.15847/obsOBS18520242445

视频：https://youtu.be/i81YFisvVEA