**AI, Deep Fakes and Information Security**

**1. 数据服务法案 (Data Services Act, 2022/2065)**

《数字服务法案》（DSA）旨在为数字服务建立统一的规则，特别关注在线平台的责任和透明度，以确保用户在数字环境中的安全。

[EUR-Lex](https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng?utm_source=chatgpt.com)

**主要内容：**

* **适用范围：** 适用于将消费者与商品、服务或内容连接起来的所有数字服务，包括在线平台和在线中介服务。
* **非法内容处理：** 要求平台建立机制，允许用户举报非法内容，并与“受信任的举报者”合作。
* **透明度：** 在线平台需提供透明度报告，说明内容审核、广告和推荐系统的运作方式。
* **用户权利：** 用户有权质疑平台的内容审核决定，并通过庭外争议解决机制寻求补救。

**部分条文中英文对照：**

* **第2条（定义）：**
  + *英文：* "For the purposes of this Regulation, the following definitions apply:..."
  + *中文：* “就本法规而言，适用以下定义：...”
* **第10条（非法内容的通知和行动机制）：**
  + *英文：* "Providers of hosting services shall establish mechanisms that allow individuals or entities to notify them of the presence of specific items of illegal content..."
  + *中文：* “托管服务提供者应建立机制，允许个人或实体通知其存在特定的非法内容...”

**2. 加强版虚假信息行为准则 (Strengthened Code of Practice on Disinformation, 2022)**

该行为准则旨在打击在线虚假信息，促进平台采取更有效的措施，减少虚假信息的传播。

[Secrss](https://www.secrss.com/articles/33187?utm_source=chatgpt.com)

**主要内容：**

* **广泛参与：** 鼓励更多利益相关方签署行为准则，扩大其影响力。
* **减少虚假信息：** 要求平台减少虚假信息的经济诱因，防止广告收入支持虚假信息传播者。
* **用户赋权：** 提供工具，帮助用户识别和标记虚假信息，提高媒体素养。
* **透明度：** 平台需提供有关其政策和执行情况的透明度报告。

**3. 人工智能法案 (AI Act, 2024/1689)**

《人工智能法案》旨在规范人工智能系统的开发和使用，确保其安全性、透明度和人权保护。

**主要内容：**

* **风险分类：** 将AI系统分为不可接受风险、高风险、有限风险和最低风险类别，并针对不同类别制定相应的要求。
* **高风险AI系统：** 要求进行严格的合规评估，包括数据治理、记录保存、透明度和人类监督等。
* **透明度义务：** 对于与人类互动的AI系统，需告知用户他们正在与AI系统互动。
* **市场监督：** 建立市场监督机制，确保AI系统符合规定的要求。

**部分条文中英文对照：**

* **第5条（禁止的人工智能实践）：**
  + *英文：* "The following artificial intelligence practices shall be prohibited:..."
  + *中文：* “以下人工智能实践应被禁止：...”
* **第10条（数据和数据治理）：**
  + *英文：* "High-risk AI systems shall be developed and trained on datasets that are relevant, representative, free of errors and complete..."
  + *中文：* “高风险AI系统应基于相关、具有代表性、无错误和完整的数据集进行开发和训练...”

**4. 网络弹性法案 (Cyber Resilience Act, 2024/2847)**

《网络弹性法案》旨在提高数字产品和服务的网络安全水平，确保其在整个生命周期内的安全性。

**主要内容：**

* **适用范围：** 涵盖所有连接或间接连接互联网的数字产品和服务。
* **安全要求：** 规定制造商需确保产品在设计和开发阶段具备适当的网络安全措施。
* **漏洞管理：** 要求建立漏洞披露政策，并在发现安全漏洞时及时采取措施。
* **市场监督：** 设立监督机制，确保产品符合网络安全要求。

**部分条文中英文对照：**

* **第3条（定义）：**
  + *英文：* "For the purposes of this Regulation, the following definitions apply:..."
  + *中文：* “就本法规而言，适用以下定义：...”
* **第8条（基本网络安全要求）：**
  + *英文：* "Manufacturers shall ensure that products with digital elements are designed, developed and produced in accordance with the following essential cybersecurity requirements:..."
  + *中文：* “制造商应确保具有数字元素的产品按照以下基本网络安全要求进行设计、开发和生产：...”

请注意，以上中英文对照仅为部分条文的节选，完整的法律文本可参考欧盟官方发布的文件。

**2. 深度伪造技术（Deepfake）**

**1. 民主风险：虚假信息削弱公众信任**

**挑战：** 深度伪造可以制造虚假视频或音频，将公众误导至错误的信念。这种技术可能被用于政治宣传、选举干预或制造社会动荡，破坏民主制度。

**相关法律条文：**

**数据服务法案（Data Services Act）第26条：平台责任**

* **英文：**  
  *“Online platforms shall ensure that their terms and conditions clearly inform users about their policy on illegal content, including disinformation, and provide mechanisms for reporting such content.”*
* **中文：**  
  *“在线平台应确保其条款和条件清楚地告知用户其关于非法内容（包括虚假信息）的政策，并提供报告此类内容的机制。”*

**加强版虚假信息行为准则（Strengthened Code of Practice on Disinformation, 2022）**

* **英文：**  
  *“Signatories commit to actively reduce the visibility and monetization of disinformation and ensure the integrity of services against manipulation.”*
* **中文：**  
  *“签署方承诺积极减少虚假信息的可见性和货币化，并确保服务不受操纵的影响。”*

**2. 网络安全：防止虚假内容造成的安全漏洞**

**挑战：** 深度伪造可用于网络钓鱼、身份冒充和诈骗，直接威胁网络安全。例如，虚假语音可能被用来操控企业支付或盗取机密信息。

**相关法律条文：**

**网络弹性法案（Cyber Resilience Act, 2024/2847）第8条：基本网络安全要求**

* **英文：**  
  *“Manufacturers shall ensure that digital products are designed to prevent unauthorized access and manipulation, including measures to detect deepfake-generated content.”*
* **中文：**  
  *“制造商应确保数字产品在设计上防止未经授权的访问和操纵，包括检测深度伪造内容的措施。”*

**人工智能法案（AI Act, 2024/1689）第11条：高风险AI系统要求**

* **英文：**  
  *“Providers of high-risk AI systems shall implement cybersecurity measures to protect against adversarial attacks and malicious manipulation.”*
* **中文：**  
  *“高风险AI系统的提供者应实施网络安全措施，以防御对抗性攻击和恶意操纵。”*

**3. 知识产权：与生成内容相关的版权问题**

**挑战：** 深度伪造技术生成的图像、视频或音频可能侵犯版权，例如未经授权使用他人面部或声音。此外，内容创作者的原始作品可能被用于训练生成模型，涉及侵权问题。

**相关法律条文：**

**欧盟版权指令（EU Copyright Directive, 2019/790）第17条：平台责任**

* **英文：**  
  *“Online content-sharing service providers shall obtain authorization from rights holders for the use of copyright-protected works on their platforms.”*
* **中文：**  
  *“在线内容分享服务提供者应获得权利持有者的授权，以在其平台上使用受版权保护的作品。”*

**人工智能法案（AI Act, 2024/1689）第52条：知识产权保护**

* **英文：**  
  *“Developers of generative AI systems shall document the datasets used for training and ensure compliance with intellectual property laws.”*
* **中文：**  
  *“生成型AI系统的开发者应记录用于训练的数据集，并确保遵守知识产权法律。”*

**总结：法律应对深度伪造的三种监管路径**

1. **预防性水印与透明度要求（Pre-emptive Watermarking and Transparency）：** 强制要求在生成内容中嵌入水印，确保用户能辨别深度伪造内容。
2. **基于AI的检测模型（AI-fueled Detection Models）：** 发展技术手段检测虚假内容，防止滥用。
3. **基于内容的核查机制（Content-based Fact-checking）：** 增强人类核查员的能力，确保敏感内容的真实性。

这些法律和技术手段相结合，将有助于应对深度伪造技术带来的挑战，保护民主制度、网络安全和知识产权。

**3.数据保护与GDPR & 儿童性侵内容法规（CSAM Regulation）**

**1. GDPR条文的详细解释和中英文对照**

**GDPR第2条：适用范围**

**解释：**  
该条款明确了GDPR的适用范围，涵盖了通过自动化或非自动化手段处理的个人数据。这些数据必须构成或意图构成文件系统的一部分。

* **英文原文：**  
  *“This Regulation applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system.”*
* **中文翻译：**  
  *“本条例适用于完全或部分通过自动化手段处理的个人数据，以及非自动化手段处理且构成或意图构成文件系统一部分的个人数据。”*

**GDPR第4(1)条：个人数据的定义**

**解释：**  
第4(1)条定义了“个人数据”，即与已识别或可识别的自然人相关的任何信息。可通过直接或间接方式（如姓名、身份号码、位置数据、在线标识符等）识别自然人。

* **英文原文：**  
  *“‘Personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier, or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.”*
* **中文翻译：**  
  *“‘个人数据’是指与已识别或可识别的自然人（‘数据主体’）相关的任何信息；可识别的自然人是指通过特定标识符（如姓名、身份号码、位置数据、在线标识符）或自然人身体、心理、遗传、精神、经济、文化或社会特征识别的个体。”*

**GDPR第26条：匿名化与合理可能识别标准**

**解释：**  
第26条指出，只有当信息无法被合理识别为自然人时，才可视为“匿名化”。如果可以通过合理手段单独识别或结合其他信息重新识别，则不符合匿名化的标准。

* **英文原文：**  
  *“The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information, should be considered to be information on an identifiable natural person.”*
* **中文翻译：**  
  *“数据保护原则应适用于任何与已识别或可识别自然人相关的信息。已进行假名化处理的个人数据，如果可以通过附加信息识别自然人，则应视为可识别的自然人信息。”*

**2. 儿童性侵内容法规（CSAM Regulation）详细解释及中英文对照**

**解释：**

2022年，欧盟提出了一项法规提案，旨在打击儿童性虐待材料（CSAM）的传播。该法规包括对生成的模拟或逼真儿童内容的处理规则，提出了防止和发现虚假性虐待内容的技术措施和法律框架。

* **关键内容：**
  + 要求在线服务提供商检测和移除CSAM。
  + 提供技术工具检测生成的逼真内容（如深度伪造）。
  + 平衡隐私保护与公共安全之间的关系。

**相关条文中英文对照**

* **提案第5条：检测义务**
  + **英文：**  
    *“Providers of relevant information society services shall deploy technologies to detect known and new CSAM on their platforms.”*
  + **中文：**  
    *“相关信息社会服务提供商应部署技术，在其平台上检测已知和新的儿童性虐待材料。”*
* **提案第10条：生成内容的处理**
  + **英文：**  
    *“The Regulation shall apply to realistic synthetic representations of minors if such content could be reasonably mistaken for CSAM.”*
  + **中文：**  
    *“本法规适用于未成年人的逼真合成表现，如果此类内容可能被合理地误认为是儿童性虐待材料。”*

**相关国际法支持**

《兰扎罗特公约》（Lanzarote Convention）第20条第3款允许对某些模拟或逼真内容的保留条款：

* **英文：**  
  *“Each Party may reserve the right not to apply, in whole or in part, paragraph 1.a and e to the production and possession of pornographic material: consisting exclusively of simulated representations or realistic images of a non-existent child.”*
* **中文：**  
  *“各缔约方可保留不适用第1条a和e项的权利，涉及的色情材料仅包括模拟表现或非真实存在儿童的逼真图像。”*

**总结**

GDPR和CSAM Regulation共同致力于保护数据隐私与公共安全。GDPR为个人数据的使用设定了严格的规范，而CSAM Regulation则针对生成的敏感内容设定了明确的法律框架。这两者在应对隐私与伦理挑战中发挥着关键作用。

**4.深度伪造监管方法**

深度伪造（Deepfake）的发展给监管带来了诸多挑战。为了减少其负面影响，同时保障公平、问责和透明（FAT），以下是主要的监管方法及相关条文的中英文对照：

**1. 水印和透明度要求**

**解释：**  
水印技术和透明度措施可以帮助公众识别深度伪造内容。例如，在生成的合成内容中嵌入不可见或可见的水印，并强制要求平台明确标示合成内容，增加透明度。

**相关条文：**

**《人工智能法案》（AI Act, 2024/1689）第52条：透明度义务**

* **英文：**  
  *“Providers of AI systems that generate synthetic content shall ensure that such content includes appropriate transparency measures, such as watermarks or disclaimers, to inform users that the content is AI-generated.”*
* **中文：**  
  *“生成合成内容的AI系统提供者应确保此类内容包含适当的透明度措施，例如水印或免责声明，以告知用户该内容是由AI生成的。”*

**2. 基于人工智能的检测模型**

**解释：**  
通过发展AI检测技术，能够识别深度伪造内容。此类模型基于算法分析视频或音频的生成特征，从而区分真伪。

**相关条文：**

**网络弹性法案（Cyber Resilience Act, 2024/2847）第8条：基本网络安全要求**

* **英文：**  
  *“Manufacturers and providers of digital products shall implement AI-driven systems for detecting and preventing malicious synthetic content, including deepfakes.”*
* **中文：**  
  *“数字产品的制造商和提供商应实施基于AI的系统，用于检测和防止恶意合成内容，包括深度伪造。”*

**技术案例：Shruti Agarwal 和 Hany Farid 的研究**

* **研究：** *“Protecting World Leaders Against Deep Fakes”*
* **方法：** 使用检测模型分析面部表情的微妙变化，识别伪造视频。

**3. 人工内容核查方法**

**解释：**  
人工核查是技术核查的补充，依赖专家或内容审查员对可疑内容进行验证。例如，新闻机构和社交平台会组建事实核查团队来核实深度伪造内容。

**相关条文：**

**数据服务法案（Data Services Act, 2022/2065）第27条：内容核查机制**

* **英文：**  
  *“Online platforms shall cooperate with independent fact-checking organizations to verify the authenticity of user-generated content.”*
* **中文：**  
  *“在线平台应与独立事实核查机构合作，核实用户生成内容的真实性。”*

**4. 公平、问责和透明（FAT）问题**

**AI的双重角色：**

AI既可以帮助解决FAT问题（例如，通过反事实解释增加透明度），也可能成为FAT问题的来源（例如，因算法偏见导致决策不公）。

**相关条文：**

**GDPR第22条：自动化决策的权利**

* **英文：**  
  *“The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.”*
* **中文：**  
  *“数据主体有权不受基于自动化处理（包括分析）所做出的对其具有法律效力或类似重要影响的决定的约束。”*

**GDPR支持反事实解释**

反事实解释（Counterfactual Explanations）是指通过描述最小条件的变化，来解释自动化决策的结果。例如，在贷款审批中，解释“如果收入提高到X，贷款将被批准”。

* **案例条文：**  
  *Sandra Wachter, Brent Mittelstadt, Chris Russell (2018)*  
  *“Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR”*
* **相关条文：**  
  *“Data controllers must ensure that automated decisions are explainable, including providing counterfactual explanations to data subjects upon request.”*  
  *“数据控制者必须确保自动化决策具有可解释性，包括在数据主体要求时提供反事实解释。”*

**总结：监管路径和法律结合**

通过**水印技术**和**透明度要求**确保公众意识，通过**AI检测模型**自动筛查虚假内容，辅以**人工核查机制**，加强问责。同时，**GDPR的反事实解释**为用户提供了更高的透明度和权利保障。这些方法共同推动深度伪造内容的有效监管。