欧盟对AI的监管方法

Gabriele Mazzini
《人工智能法案》架构师和主要作者
MIT Fellow

第十四次GIO圆桌会议
上海文华东方酒店
2024年9月20日

欧盟机构架构/流程

立法

委员会提案：2021年4月21日

欧盟理事会
一般方法：2022年12月6日

议会立场：
2023年6月14日

2024年7月12日：在欧盟官方公报上发布
2024年8月1日：正式生效

2022年11月30日：OpenAI ChatGPT上线

2023年3月22日：呼吁暂停AI训练/研究比GPT-4更强大的系统

2023年10月30日：《关于安全、可靠、可信地开发和使用人工智能的行政命令》

实施

欧盟成员国

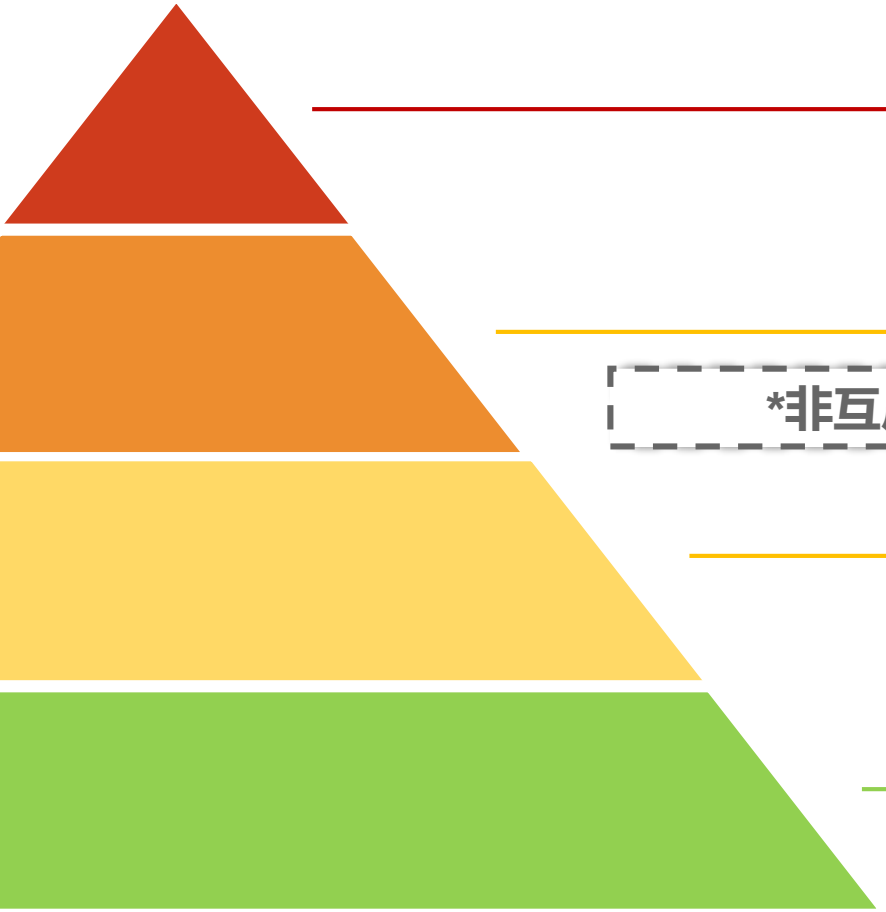
《人工智能法案》的本质

- **为人工智能系统（带CE标志）的上市和投入使用设立“基本的”内部市场规则**
- **确保将健康和安全以及其他基本权利**视为受法律保护的权益并为其提供**高水平的保护**
- **水平式管理方法**：欧盟职权范围内的跨部门管理。不涉及国家安全，军事，以及国防
  - 考虑的部门特性/需求（执法、金融和现行的产品立法）
  - 在不损害欧盟其他现行的相关法律（例如数据保护法规、消费者保护指令、与促进平等有关的法律、针对平台的立法）的情况下：*《人工智能法案》并非唯一适用于AI的欧盟法律*

基于风险的方法

风险越高，规则越严格

没有对技术本身进行监管



*非互斥

不可接受风险
如：社会信用评分

禁止项

高风险
如：招聘和医疗器械

在符合AI要求和通过事前符合性评估的前提下可**允许**

“透明度”风险
“模仿”（机器人）和深度伪造

允许，但须履行信息/透明度义务

风险极小或无风险

无限制条件的**允许**

通用人工智能（GPAI）模型

所有GPAI
（较低层级）

- 技术文档（包括计算资源和能耗）
- 信息下游
- 版权（政策以及详细内容总结）

存在系统性风险的GPAI
（较高层级）

- 对高影响力能力的评估
  - 每秒执行的浮点运算次数（FLOPS）至少为10^{25}次
  - 由AI办公室指定（例如基于某些标准）
- 较低层级的所有义务加上
  - 风险评估和规避
  - 事件报告
  - 适当的网络安全水平

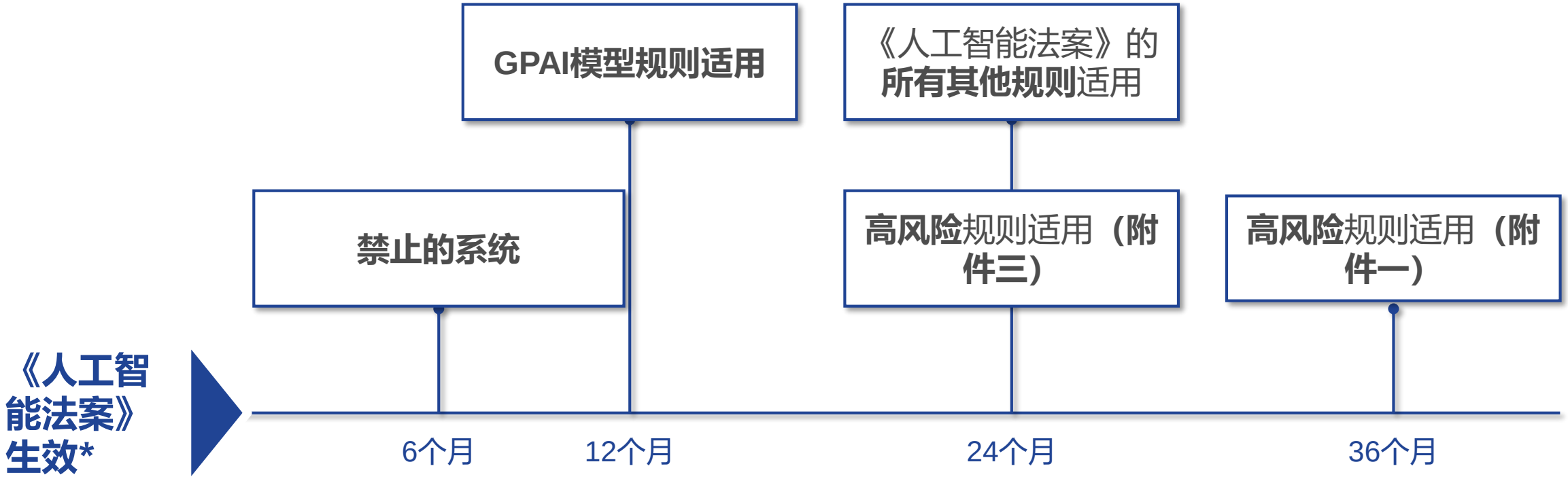
- 范围内的**开源模型**，较低层级的技术文档和透明度除外
- 用于证明合规性的**实践守则**

AI办公室与实施

《人工智能法案》的实施

- 执行GPAI规则（包括开发评估工具/基准和调查侵权行为）
- 支持国家主管部门执法
- 支持委员会履行其所有职责（例如提供指导、授权/实施法案[约70个行动项目]、作为AI理事会及其小组秘书处，以及为咨询论坛和科学专家小组提供行政支持）
  - 第96条（应）：适用第8至15条和第25条所述的**要求和义务**；第5条所述的**禁止做法**；实际执行与**重大修改**有关的规定；实际执行第50条规定的**透明度义务**；关于**本规定与附件一所列欧盟协调立法以及与其他相关欧盟法律的关系**的详细信息，包括这些法律法规在执行方面的一致性；第3条第（1）点中规定的**AI系统定义的适用情况。**

逐步应用



谢 谢!

gmazzini@llm10.law.harvard.edu