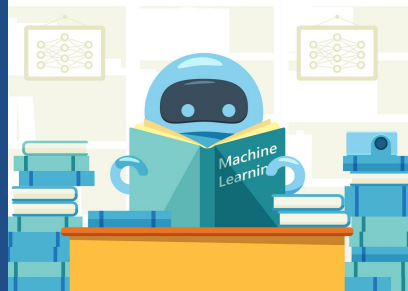


AI

Machine Learning

机器学习



生成式对抗网络

节选自 第四章 - 神经网络



软件工程专业 三年级 专业选修课



3学分 64学时
(含32上机学时)

导入 - 为电影角色“换脸”



小罗伯特·唐尼

汤姆·克鲁斯



钢铁侠

生成式人工智能模型，能够产生足够“以假乱真”的数据

视频、图片来源于网络

生成式对抗网络的基本概念



- 生成对抗网络是一种生成模型

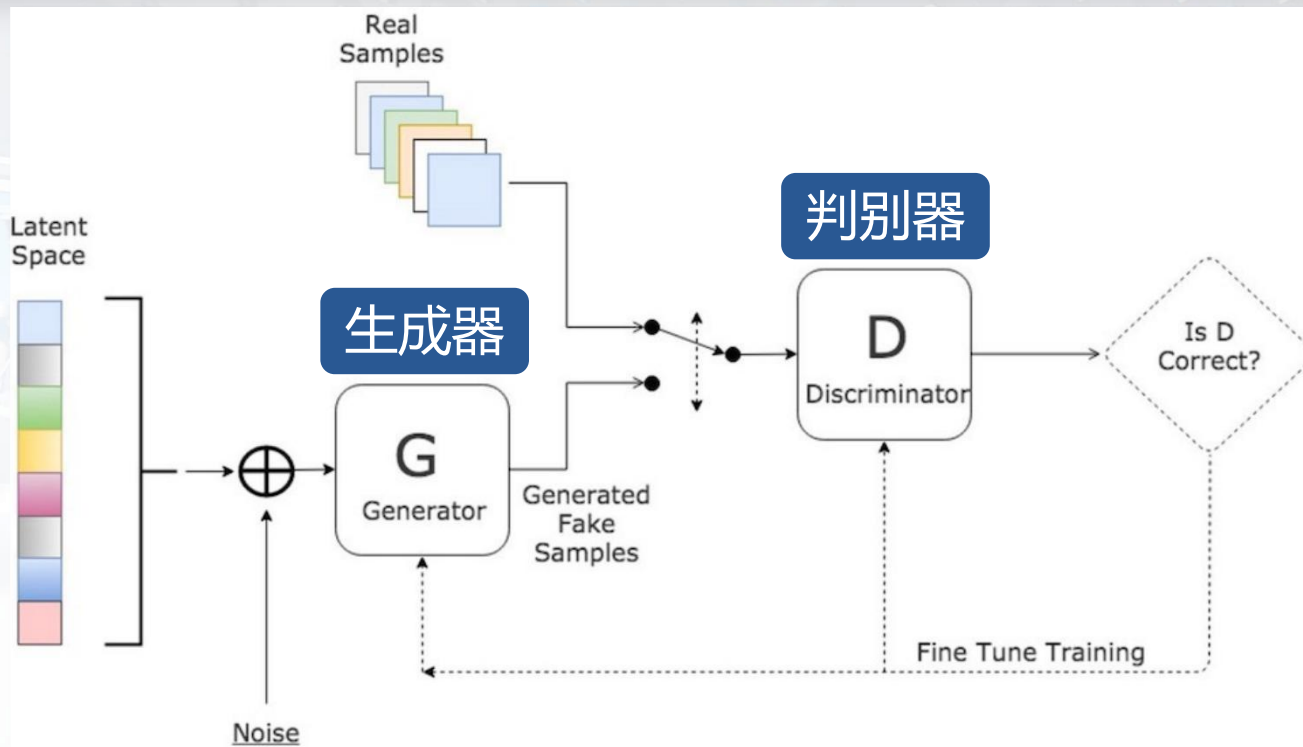
Generative Adversarial Networks, 简称为 GAN, 读作 “gàn”

- 什么是生成模型呢？

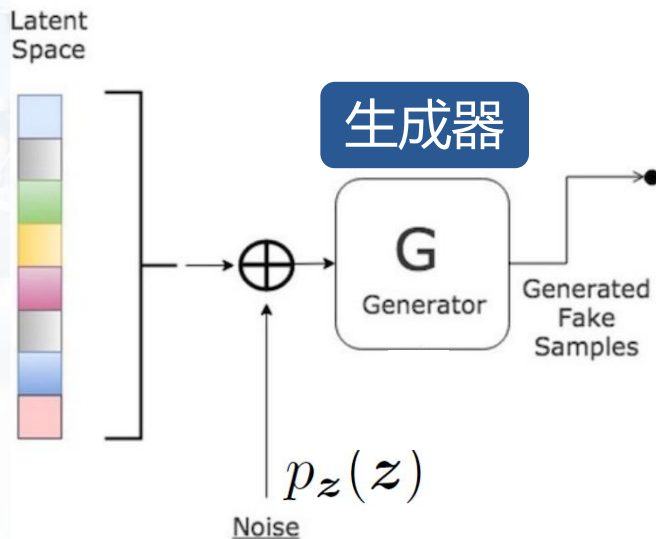


在机器学习中：生成模型可以用来直接对数据建模，也可以用来建立变量间的条件概率分布

生成式对抗网络的架构



生成式对抗网络的架构

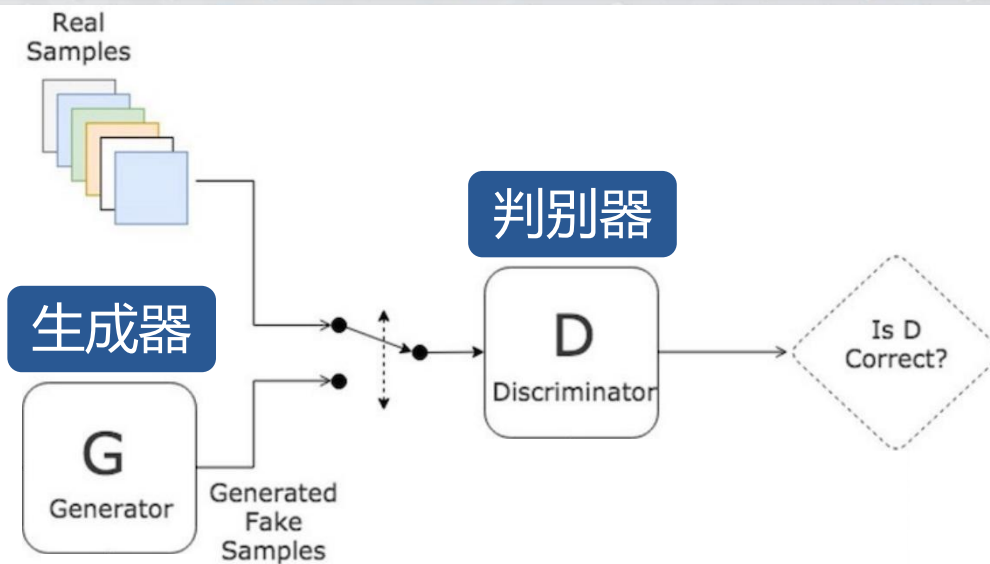


生成器

θ_g

$G(z; \theta_g)$

生成式对抗网络的架构



生成器

θ_g

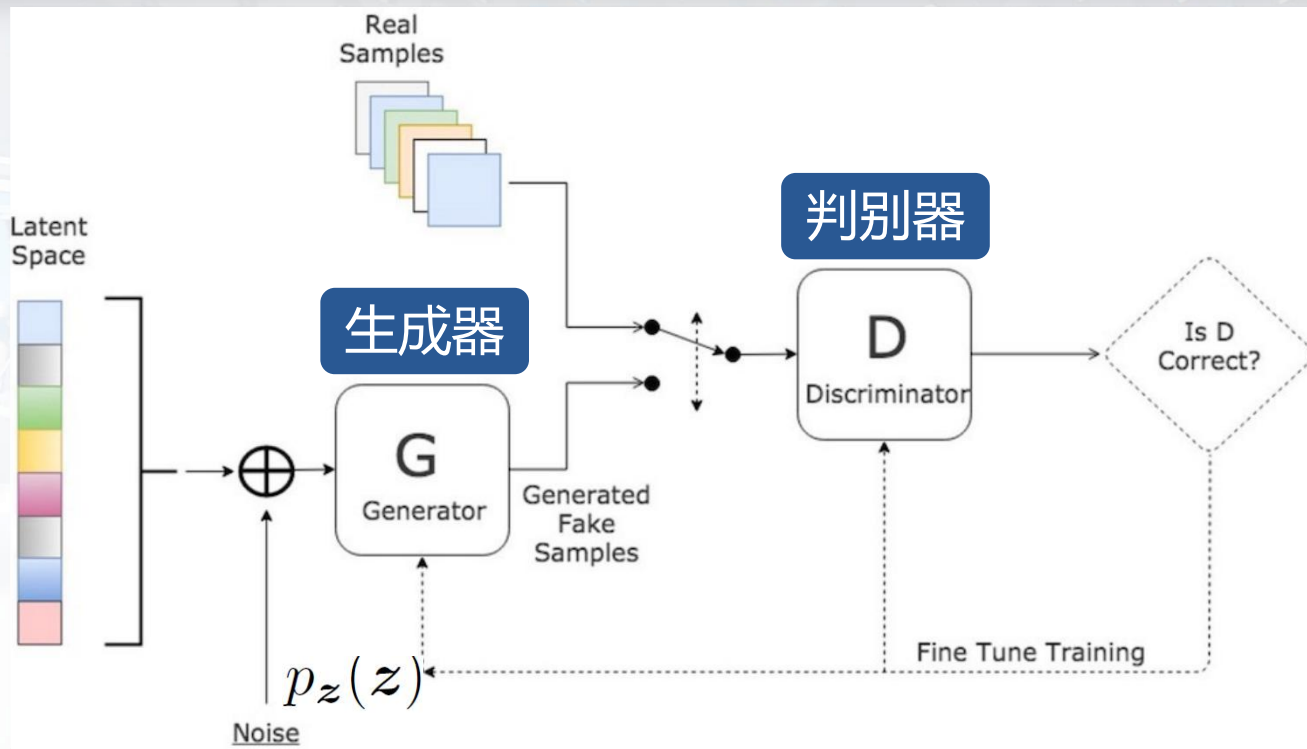
$G(z; \theta_g)$

判别器

θ_d

$D(x; \theta_d)$

生成式对抗网络的架构



生成器

θ_g

$G(z; \theta_g)$

判别器

θ_d

$D(x; \theta_d)$

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log(D(x))] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

生成式对抗网络的优化

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log(D(x))] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$



优化方式:

- 生成器的优化
- 判别器的优化

生成式对抗网络的优化 - G

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log(D(x))] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$



将D的参数都看成常数

$$\min_G V(D, G) = E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$



生成式对抗网络的优化 - G

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log(D(x))] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

将D的参数都看成常数

$$\min_G V(D, G) = E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

梯度下降法

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(z^{(i)})))$$

生成式对抗网络的优化 - D

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log(D(x))] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

↓
将G的参数都看成常数

$$\max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

↓

生成式对抗网络的优化 - D

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log(D(x))] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

将G的参数都看成常数

$$\max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

梯度下降法 $\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log D(\mathbf{x}^{(i)}) + \log(1 - D(G(\mathbf{z}^{(i)})))]$

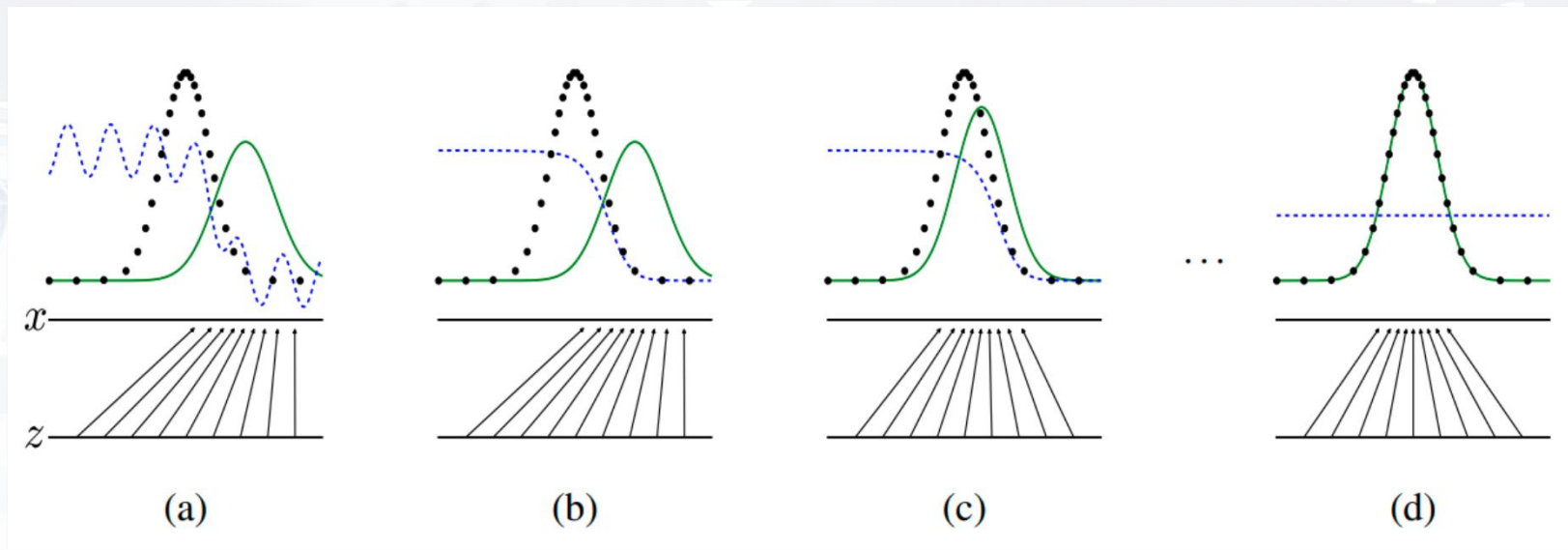
生成式对抗网络的优化

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log(D(x))] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

优化方式：

- 生成器的优化
- 判别器的优化
- 交替优化，直到达到纳什均衡点（若在某种情况下无一参与者可以独自行动而增加收益，则此策略被称为纳什平衡点）

生成式对抗网络的优化

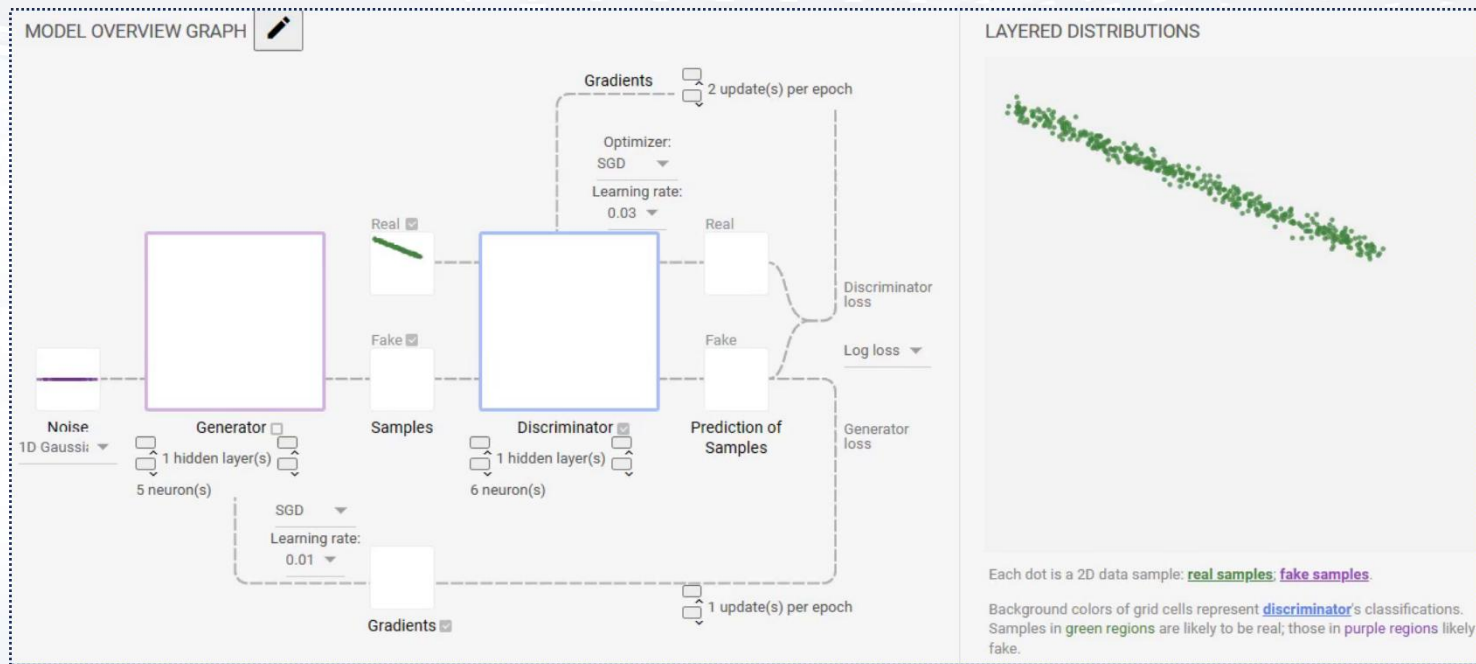


z —高斯噪声的域 x —真实样本的域 黑色点—真实样本的分布 蓝色点—分类器分类的表现 绿色线—生成数据的分布

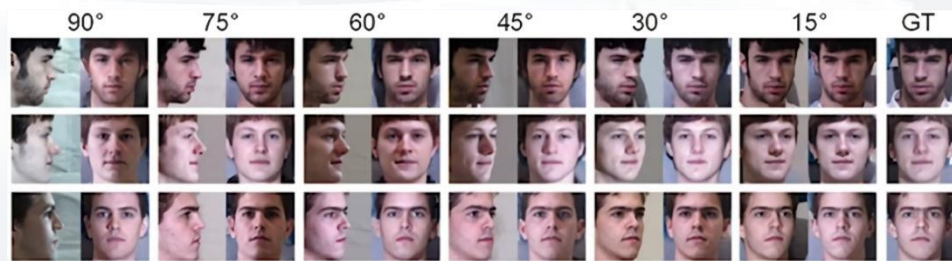
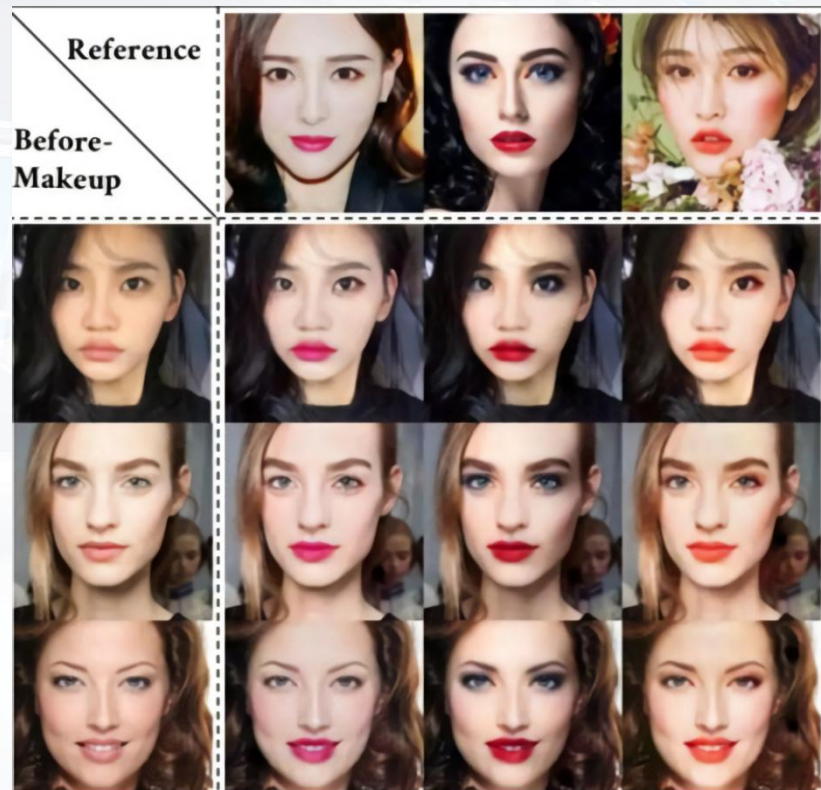
最终，生成模型生成的数据和真实样本一致，判别器 D 无法准确划分， $D(x) = \frac{1}{2}$

生成式对抗网络的优化

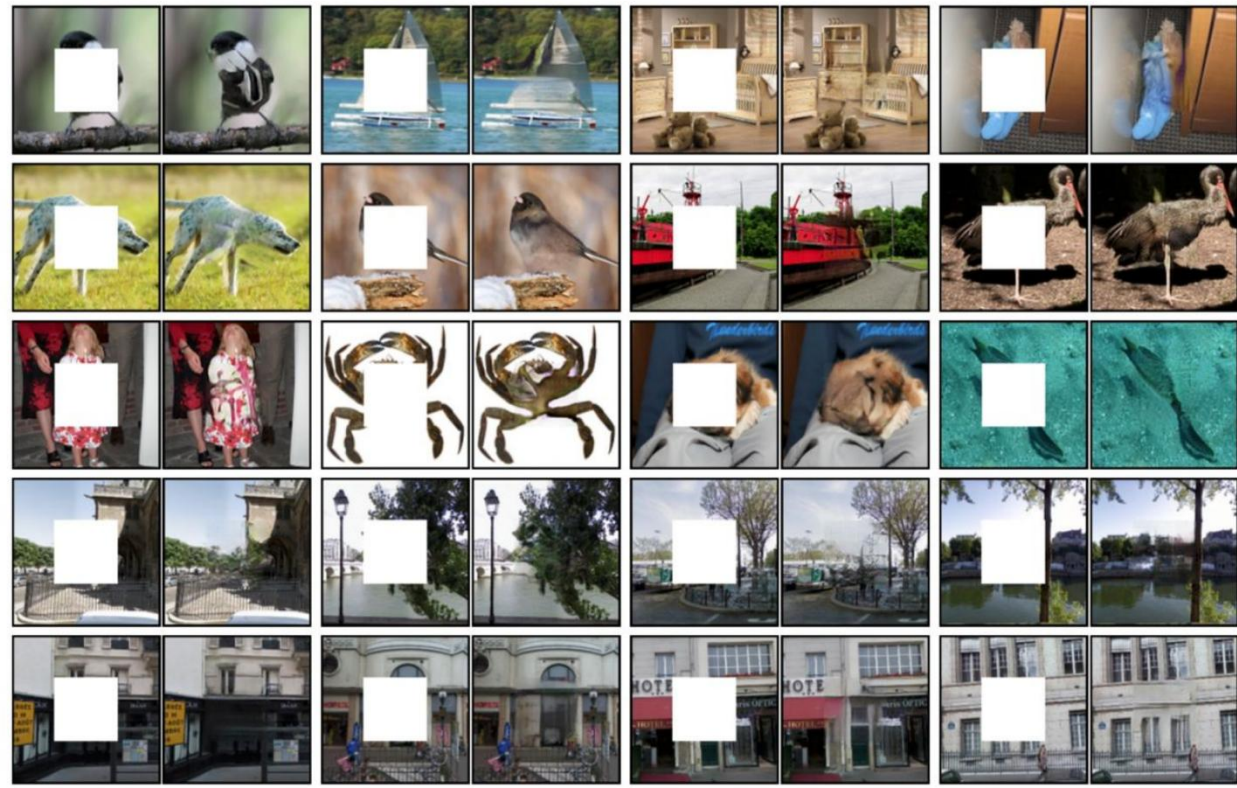
$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log(D(x))] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$



生成式人工智能应用-人脸图像



生成式人工智能应用-图像修复



生成式人工智能应用-生成视频



中新网3月24日电 “特朗普被捕”的事情还没有发生，但“特朗普被捕”的“照片”已经在一些社交平台流传开来。



Eliot Higgins

@EliotHiggins

Making pictures of Trump getting arrested while waiting for Trump's arrest.



假



假



假

生成式人工智能

公开征求意见系统

欢迎进入公开征求意见系统！感谢您的支持和参与！

法律法规草案(0部)

法律法规草案名称	征求意见单位	草案说明	起止时间
----------	--------	------	------

部门规章草案(9部)

部门规章草案名称	征求意见单位	草案说明	起止时间
住房和城乡建设部关于修改《建设工程勘察设计审查办法》	住房和城乡建设部		2023-04-20 至 2023-05-20
上市公司独立董事管理办法（征求意见稿）	证监会	说明	2023-04-14 至 2023-05-14
期货从业人员管理办法（征求意见稿）	证监会	说明	2023-04-14 至 2023-05-14
生成式人工智能服务管理办法（征求意见稿）	国家互联网信息办公室		2023-04-11 至 2023-05-10
矿山救援规程（修订草案）	应急管理部	说明	2023-04-06 至 2023-05-06
期货公司监督管理办法（征求意见稿）	证监会	说明	2023-03-24 至 2023-04-23
宗教活动场所管理办法(征求意见稿)	宗教事务局		2023-03-24 至 2023-04-24
健身气功管理办法（修订草案）	体育总局	说明	2023-03-20 至 2023-04-21
航空体育运动管理办法（修订草案）	体育总局	说明	2023-03-20 至 2023-04-21

征求意见已结束的
法律法规草案

征求意见已结束的
部门规章草案

中国法律服务网 (12348中国法网)
Public Legal Services of China
中文域名: 12348中国法网 公益

中华人民共和国司法部 版权所有 国京ICP备0505517 8号
司法部信息中心 运维保障 Copyright © 2017 All Rights Reserved

政府网站
找错



中华人民共和国国家互联网信息办公室

Cyberspace Administration of China

WWW.CAC.GOV.CN

请输入检索关键词

首页

权威发布

办公工作

网络安全

信息化

网络传播

国际交流

地方网信

执法监督

政策法规

互动中心

教育培训

业界动态

工作专题

当前位置: 首页 > 正文

国家互联网信息办公室关于《生成式人工智能服务管理办法（征求意见稿）》公开征求意见的通知

2023年04月11日 12:51 来源: 中国网信网

为促进生成式人工智能技术健康发展和规范应用,根据《中华人民共和国网络安全法》等法律法规,国家互联网信息办公室起草了《生成式人工智能服务管理办法(征求意见稿)》,现向社会公开征求意见。公众可通过以下途径和方式提出反馈意见:

- 1.登录中华人民共和国司法部 中国政府法制信息网 (www.moj.gov.cn、www.chinalaw.gov.cn)、进入首页主菜单的“立法意见征集”栏目提出意见。
- 2.通过电子邮件方式发送至: wqjcyj@cac.gov.cn。
- 3.通过信函方式将意见寄至: 北京市海淀区阜成路15号国家互联网信息办公室网络安全协调局,邮编100048,并在信封上注明“生成式人工智能服务管理办法征求意见”。

意见反馈截止时间为2023年5月10日。

附件: 生成式人工智能服务管理办法(征求意见稿)

国家互联网信息办公室
2023年4月11日

为促进生成式人工智能健康发展和规范应用，国家互联网信息办公室，制定了《生成式人工智能服务管理办法》部门规章草案

第四条

提供生成式人工智能产品或服务应当遵守法律法规的要求，尊重社会公德、公序良俗

第十二条

提供者不得根据用户的种族、国别、性别等进行带有歧视性的内容生成。

节选自国家网信办《生成式人工智能服务管理办法》征求意见稿