Machine Translated by Google

第25话

盘点

但是:连接世界意味着我们也连接了所有坏事和坏人,现在每个社会和政治问题都用软件来表达。我们经历了一个可怕的“糟糕”的意识到时刻,但我们还没有远程想出该怎么办。

– 本尼迪克特埃文斯

如果你为自由而战,你很可能会发现自己在酒吧的错误一端和坏人一起喝酒。

– 惠特死

25.1 简介

我们在剑桥的安全小组运营着一个博客 www.lightbluetouchpaper.org,我们在其中讨论最新的黑客攻击和破解。许多攻击都取决于特定的应用程序,许多很酷的研究也是如此。但是,并非所有应用程序都是相同的。如果我们的博客软件被黑客攻击,它只会给僵尸网络多一台服务器,但还有其他应用程序可以窃取金钱,其他应用程序是人们赖以保护隐私的应用程序,其他应用程序可以调解权力,还有一些可以杀人。

我已经讨论过许多应用程序,从银行业务到警报再到预付费电表。在本章中,我将简要描述处于安全研究前沿的四类应用程序。它们是我们发现创新攻击、新保护问题和棘手政策问题的地方。它们是: 自动驾驶和遥控车辆;机器学习,从对抗性学习到人工智能在社会中更普遍的问题;隐私技术;最后,电子选举。它们的共同点是,虽然以前安全工程是关于管理技术的复杂性及其所有可利用的副作用 ,但我们现在正在应对人类社会的复杂性。自动驾驶汽车很难,因为人们在同一条道路上驾驶其他汽车。人工智能很难,因为我们很酷的新模式匹配工具,例如深度神经网络,不仅可以挑选出人类行为中的真实模式

Machine Translated by Google

有时是意想不到的 但也是错误的。由于社会中人类互动的丰富性,隐私很难实现。选举之所以艰难,不仅是因为以保护隐私和可审计性的方式计票存在技术难度,还因为政治参与者在投票过程本身的上游和下游使用了各种各样的肮脏伎俩。所有这些问题都以不同的方式探索了人类能做什么和机器能做什么之间的界限。

25.2 自主和遥控驾驶车辆

航空先驱劳伦斯·斯佩里 (Lawrence Sperry) 于 1912 年发明了第一台自动驾驶仪,并于 1914 年进行了演示,在巴黎举行的“更安全的飞机”比赛中,他举起双手飞越了评委。在此过程中,他和他的父亲埃尔默发明了人造地平线。一架自行其是的固定翼飞机最终会进入螺旋俯冲并坠毁;飞行员可以参考地平线保持它的水平,但是当在云中飞行时,外部参考丢失了。陀螺仪可以提供缺失的参考,它还可以通过舵机驱动副翼和升降舵。

1975 年,我找到了第一份正式工作,重新设计了一个快速喷气式惯性导航装置,用于石油工业中使用的小型潜艇。同一栋楼里的工程师正在研究早期的平视显示器和卫星导航设备。这些设备每件重约 20 公斤,价值 250,000 英镑 相当于今天的 300 万美元。这三个人加起来一千万美元,几乎没有零钱,而且和一个人一样重。

现在,在 2020 年,您的手机中同时具备了这三种功能。您的手机不是在精密设计的笼子中安装三个旋转的机械陀螺仪,而是带有带有 MEMS 加速度计和陀螺仪的芯片。它还有一个用于卫星导航的 GPS 芯片和一个谷歌或苹果地图应用程序来告诉你如何步行、骑自行车或开车到你的目的地。四十多年来,成本下降了六个数量级,质量下降了四个数量级。这推动了海上、空中和陆地辅助技术的快速发展。 1920 年代开创性的单人帆船运动员开发了自舵装置来横渡大洋,让他们有时间睡觉、做饭和修理帆;业余爱好者现在拥有更智能的沿海巡航自动驾驶仪。自主探测器在南极冰层下方游动,以测量其融化的速度。世界各国的海军都在开发水雷、用于寻找水雷的自主潜水器等等。

25.2.1 无人机

在空中,德国 V1 和 V2 等早期武器使用了双陀螺仪自动驾驶仪,而冷战则为我们提供了在两次海湾战争中都发挥了巨大作用的战斧巡航导弹。自 20 世纪 80 年代初期开始服役,这些飞机通过贴近地面飞行潜入敌方雷达之下,并使用地形等高线匹配来更新其惯性导航。紧随其后的是各种无人驾驶飞行器 (UAV),这些飞行器于 1982 年在以色列和叙利亚之间的战争中首次大规模使用;以色列空军将它们用作侦察和诱饵,以最小的损失消灭了叙利亚空军。这

Machine Translated by Google

最著名的下一代无人机是“捕食者”。最初设计为侦察车,它可以在中等高度的目标区域上空逗留数小时,并适合携带地狱火导弹打击地面目标。从 1995 年到 2018 年服役,它在伊拉克、阿富汗、利比亚和其他地方服役。它被更大、更快的死神所取代,后者成为叙利亚打击伊斯兰国战争的中流砥柱。世界武装部队现在拥有范围广泛的无人机,小到士兵们在背包中携带的小型无人机,它们可以用来查看下一个拐角处的情况。

整个 20 世纪,爱好者们都在建造小型无线电遥控模型飞机,但美国联邦航空局 (FAA) 直到 2006 年才颁发了第一份商用无人机许可证。
2010 年,Parrot 推出了 AR 无人机,这是一款可以通过智能手机 wifi 控制的四轴飞行器,2013 年,亚马逊宣布正在考虑使用无人机进行送货。兴趣迅速升起;几年之内,我们的学生就开始制造无人机,很快你就可以在业余爱好商店买到低成本的型号。 2020年的主要应用是航拍。不过,无人机既有叛乱分子用途,也有犯罪用途,无人机曾被用来向囚犯运送毒品和手机,而叛乱分子则为无人机安装了简易爆炸装置,用作武器。

25.2.2 自动驾驶汽车

不过,最近激增的兴趣大多集中在自动驾驶汽车和卡车上。 2004 年,面对阿富汗和伊拉克的简易爆炸装置不断增加的战斗损失,DARPA 决定推动自动驾驶汽车的发展,并宣布了一场比赛,谁能制造出能够穿越 149 英里公路的汽车,谁就能获得百万美元的奖金。莫哈韦沙漠最快。

由于没有车辆完成该课程,该奖项无人认领,但第二年,由机器人专家塞巴斯蒂安·特伦 (Sebastian Thrun) 领导的斯坦福大学团队获得了 200 万美元的奖金。他的机器人 Stanley 使用机器学习和概率推理来处理地形感知、避免碰撞以及在湿滑和崎岖地形上稳定车辆控制 [1887]。这建立在可追溯到 80 年代的机器人研究的基础上,其中大部分研究也得到了 DARPA 的资助。他们在 2007 年的下一个挑战是从沙漠转移到模拟城市环境;参赛者必须发现并避开其他车辆,并遵守道路规则。这引导了一个研究社区,技术开始迅速改进。

以前,汽车制造商一直在稳步增加辅助技术,从上个世纪的防抱死制动系统 (ABS) 开始,然后发展到自适应巡航控制 (ACC),我在第 23.4.1 节中描述了自动紧急制动 (AEB) 和车道保持协助（LKA）。行业的愿景是,这些最终将组合成一个全自动驾驶仪。受 DARPA 挑战的启发,谷歌于 2009 年聘请塞巴斯蒂安·特伦 (Sebastian Thrun) 领导 Project Chaueur,目标是打造一款全自动驾驶汽车。这是在 2010 年宣布的,刺激了一场涉及科技和汽车行业的市场竞争。

特斯拉于 2014 年率先推出一款产品,当时其“Autopilot”软件作为无线升级推出,可以控制高速公路或起停交通。机器学习的炒作周期已经开始,我将在下一节中讨论,自动驾驶汽车也搭上了顺风车

Machine Translated by Google

骑。在谷歌汽车项目于 2016 年作为 Waymo 分拆之前,特斯拉的埃隆马斯克预测到 2018 年实现完全自主,谷歌的谢尔盖布林则预测到 2017 年实现完全自主。 -服务； Uber 的到来又增加了一个竞争对手,这种炒作甚至吓坏了汽车行业的高管,他们应该更清楚地预测到 2020 年代中期人们将不再拥有自己的汽车。炒作周期一如既往地过去了。正如我在 2020 年写的那样,Waymo 正在凤凰城 50 平方英里的区域内运营有限的自动驾驶汽车服务 [871]。该服务在下雨或沙尘暴时不可用,并由控制中心的人员实时监控。它已经宣布了好几次,但问题一直存在,迫使公司将安全驾驶员放回车内。发生什么了？

很大一部分答案是其他道路使用者是不可预测的。自动化可以处理由此产生的一些危险:如果前面的车突然刹车,机器人可以更快地做出反应。一旦有足够多的车辆使用自适应巡航控制系统,它就可以减少驾驶员疲劳,甚至可以减少拥堵,因为它可以抑制冲击波在道路上的传播。但即使在这里也有限制。当工程师将该技术扩展到自动紧急制动时,无法推断其他驾驶员的意图成为一个限制因素。例如,假设您正在开阔的乡间公路上行驶,这时前面的汽车指示转弯并开始减速。您按照预期保持速度,当您到达那里时它会离开道路,否则您就会超车。但 AEB 可能不理解这一点,所以当你离转弯的汽车太近时它就会启动,系好安全带将你抛向前方。 2020 年消费者对 AEB 系统的测试仍然显示出相当大的可变性,无论是在误报率还是在假人被拉过马路时及时停车的能力方面。一些系统将激活限制在城市而不是高速公路速度,并且在 2020 年,所有系统都倾向于在更昂贵的模型上提供选项。 AEB 应该会在 2022 年左右出现在所有新车中。自 2016 年以来,保险公司一直很高兴它降低了整体风险;我将在 28.4.1 节中讨论安全保证。

但是每一项新的辅助技术都需要数年的时间来优化和调试,而且将其中的一打组合成一个自动驾驶仪并不容易。 Sebastian Thrun 和他的团队撰写的描述 Stanley 的论文提供了对整体技术的有用见解 [1887]。有几十个程序松散地交互,反映了我们对人类如何执行此类任务的理解;你的潜意识会观察各种各样的事物,并将危险引起你的注意。斯坦利的同步流程处理路径规划、转向控制和避障;这使用了高达 22 米的激光测距仪,一个超过 22 米的彩色相机,以及一个超过 22 米的雷达（在比赛中没有使用,因为斯坦利在预定的路线上有超过 2000 个航路点）。这些系统中的每一个都必须解决许多子问题;例如,视觉系统必须适应不断变化的光照条件和道路颜色。然后必须通过大量测试对 Stanley 进行优化,其中目标函数是最大化灾难性故障（定义为人类安全驾驶员接管)之间的平均距离。

组合子系统意味着妥协,虽然主要供应商对他们的设计细节保密,但我们开始了解优化以及它们因事故而出现的问题。例如,当一个自

Machine Translated by Google

2018 年 3 月,在亚利桑那州驾驶 Uber 撞死了 Elaine Herzberg,在 NTSB 调查中发现 Elaine 一直在推自行车,视觉系统在将她识别为行人和其他人之间摇摆不定,但最终她没有被识别为行人因为她不在人行横道上。 AEB 可能已经停止了汽车,但它已被关闭 “以减少车辆行为不稳定的可能性” 换句话说,因为误报率很烦人 [457]。最终,优步依赖于安全司机 不幸的是,他当时正在看电视1。

几十年来,我们已经知道在紧急情况下依靠人类接管需要时间:人类必须对警报做出反应,分析控制台上的警报显示,扫描环境,获得态势感知,进入光流,并采取有效的控制。即使在商业航空中,自动驾驶仪发生故障后,机组人员也需要大约八秒钟的时间才能正确重新获得控制权。您不能指望汽车中的安全驾驶员会做得更好。

25.2.3 自动化的水平和限制

出于这些原因,汽车工程师协会制定了五个自动化级别:

1. 驾驶辅助 软件控制转向或速度,以及
   人类司机完成剩下的工作;
2. 部分自动化 软件在某些模式下同时控制转向和速度,但人类驾驶员负责监控环境,并在软件出现混乱时零通知地接管控制权;
3. 有条件的自动化 软件监控环境,控制转向和速度,但假设人类在感到困惑时可以接管;
4. 高度自动化 软件监控环境并在某些驾驶条件下驾驶汽车,而无需假设人类可以进行干预。如果它感到困惑,它会停在路边;
5. 完全自动化 软件可以做人类能做的一切。

到目前为止,大众市场上的车辆只有高级驾驶员辅助系统 (ADAS),即一级和二级,保险公司认为“自主”和“自动驾驶”等词是危险的,因为它们会让客户认为车辆是在 4 级运行,这可能导致事故。

亚利桑那州的撞车事故可以看作是汽车在 2 级运行,而安全驾驶员在 3 级运行。4 级通常假设有一名备用驾驶员坐在控制中心,监督几十辆“自动”汽车,但他们没有带宽以像现场的安全驾驶员一样快速了解危险。他们感觉不到道路噪音和加速度,无法使用周边视觉,最重要的是,他们没有沉浸在光流场中

1 事实上,特斯拉在自动驾驶时发生的第一起致命车祸夺去了一名司机的生命
当他的车在卡车下行驶时,他似乎正在用笔记本电脑看电影 [1394]。

Machine Translated by Google

正如我们在 3.2.1 节中讨论的那样,这对于安全驾驶汽车（或降落飞机)至关重要。

除非我们发明通用人工智能,否则 5 级在多大程度上可行?约翰·诺顿 (John Naughton) 表示,市区送货司机的工作非常安全,因为这项工作需要进行各种判断,例如,您是否可以并排停车,甚至可以在狭窄的街道上停半分钟,同时冲到门口放下一辆包裹,因为后面的汽车对你按喇叭 [1417]。

另一个棘手的案例是杂乱无章的郊区街道,两边都停着汽车,你永远在谈判谁先驶向迎面而来的车辆,使用挥手、点头甚至只是眼神交流。即使是当前的 2 级系统,由于无法进行这种默契协商,在跨越 trac 时也往往会遇到困难。他们最终不得不比人类司机更加谨慎,等待更大的差距,这惹恼了他们身后的人类司机。而且,如果您曾在剑桥这样的大学城或印度的任何城市交通中尝试让汽车从成群结队的骑自行车的学生中轻松通过,您就会知道在许多其他情况下处理复杂的人类交通是很困难的。你的自动驾驶汽车甚至能检测到警察发出的停止手势信号,更不用说应付八名抬床的学生,或应付印度寺庙游行了吗?

截至 2020 年,二级系统有很多缺点。特斯拉不能总是可靠地检测到静止的车辆;它使用视觉、声纳和雷达,但没有激光雷达。
（北卡罗来纳州的一名特斯拉司机在撞上一辆静止的警车 [1118] 后被起诉。）路虎揽胜无法始终检测到铺砌道路和草地之间的边界,但也许这不是优先事项对于 4 x 4。许多汽车都存在小回旋处的问题,更不用说坑洼和其他粗糙的表面了;我第一次坐其中,当我们以将近 30 英里/小时的速度通过减速带时,我的牙齿嘎嘎作响。道路工程对自动车道保持系统造成严重破坏,因为被涂上的旧白线可能会变成闪亮的黑色,并且在某些光线条件下非常突出,导致汽车在新旧标记之间来回摆动 [632]。有大量关于此类技术主题的研究,从更好的多传感器数据融合算法到可以为他们的决定提供解释的驾驶算法,再到让汽车在行驶时学习路线,就像人类一样。特斯拉甚至为其自动驾驶仪配备了“影子模式”;当它不被使用时,它仍然会尝试预测司机接下来会做什么,并记录它的错误预测以供以后分析。这使特斯拉能够在广泛的道路和天气条件下收集数十亿英里的训练数据。

我将在第 28.4.1 节中讨论安全保证,但 2020 年的情况是,虽然特斯拉和 NHTSA 声称特斯拉客户激活自动转向后撞车事故较少,但一个独立实验室声称撞车事故更多。正如我在第 14.3.1 节中讨论的那样,开车时睡着是事故的主要原因,占英国事故总数的 20%。这些往往处于严重的范围内;它们约占致命事故的 30%,占高速公路致命事故的一半。 （这就是为什么我们有法律限制商业司机的工作时间。）所以我们应该能够通过一个系统来挽救生命,这个系统可以让你的车在高速公路上保持在车道上,刹车以避免碰撞,并让它停在路边如果您不响应提示音,则会影响道路。为什么这没有发生?

我怀疑我们至少需要理清三个不同的因素:风险

Machine Translated by Google

恒温器、系统的规定以及市场营销产生的期望。首先,风险恒温器是人们通过采取更多风险行为来适应风险降低的机制;我们在第 3.2.5.7 节中指出,强制性安全带法律导致人们开得更快,因此总体效果只是将伤亡人员从车辆乘员转移到行人和骑自行车的人身上,而不是减少他们的总体人数。其次,正如我们在第 3.2.1 节中讨论的那样,法令规定了我们如何与技术互动,如果驾驶员辅助系统使驾驶变得更容易,并且显然更安全,人们就会放松并认为它更安全 将其中一些人置于承担更多的风险。第三,行业营销在潜移默化中将风险降到最低。

特斯拉称其自动驾驶功能为自动驾驶仪,误导司机认为他们可以看电视或小睡。飞机上的自动驾驶仪并非如此,但大多数非飞行员不明白这一点。

25.2.4 如何破解自动驾驶汽车

道路车辆的电子安全始于上个世纪,我们在 14.3 节中讨论了卡车行驶记录仪和限速器,以及我们在 4.3.1 节中讨论的远程钥匙输入系统。自 2005 年左右以来,它已成为一门专业学科,当时汽车制造商和一级零部件供应商开始聘请专家。到 2008 年,人们开始研究发动机控制单元的防篡改功能:该行业已经开始使用软件来控制发动机功率输出,因此无论您的汽车是 120 马力还是 150 马力,都取决于人们自然会尝试破解的软件开关。制造商试图阻止他们。他们声称他们担心改装不当的汽车对环境的影响,但如果你相信这一点,我有一座桥我想卖给你。

2010 年,卡尔·科舍尔 (Karl Koscher) 及其同事展示了如何破解一辆新型福特汽车,从而引起了学术界的关注。汽车的内部数据通信使用没有强认证的 CAN 总线,因此控制(比如)无线电的攻击者可以升级此访问权限以操作门锁和刹车 [1085]。 2015 年,查理·米勒 (Charlie Miller) 和克里斯·瓦拉塞克 (Chris Valasek) 入侵了一辆载有志愿记者的吉普切诺基 (Jeep Cherokee),通过其手机链接,使车辆减速并驶离道路 [1316],从而引起了媒体的注意。这迫使克莱斯勒为软件补丁召回 140 万辆汽车,公司损失超过 10 亿美元。这终于引起了业界的关注。

现在有各种各样的人对汽车和其他车辆进行黑客攻击。
有些爱好者想要调整他们的汽车;有的车库也想使用第三方组件和服务;正如我在第 24.6 节中提到的,尽管约翰迪尔垄断了服务,但仍有农民想要修理他们的拖拉机。有开源软件活动家和安全倡导者相信,如果一切都被记录下来,我们都会更安全 [1792]。还有黑帽子:想要监视车内人员的情报机构和只想偷车的小偷。

汽车盗窃是目前的主要威胁模型,我们在 4.3.1 节中讨论了用于破坏远程钥匙输入和警报系统的方法。国家行为体和其他人可以使用 2.2.1 节中讨论的技术接管嵌入汽车的移动电话。电话、导航和信息娱乐

Machine Translated by Google

不管怎样,系统通常设计得很糟糕 当你租车或买二手车时,你经常会看到以前用户的个人信息,我们在第 22.3.3 节中描述了一个应用程序如何让你跟踪和解锁租车让一旦汽车被租给别人,你就继续这样做。

那么还有什么可能会出问题,尤其是随着汽车变得更加自主?

一个合理的最坏情况可能是一个国家行为者,或者一个环境活动家团体,试图通过同时造成数千起道路交通事故来恐吓公众。诸如克莱斯勒吉普车上的远程攻击可能已经做到了这一点。大多数现代汽车用于内部数据通信的 CAN 总线信任其所有节点。如果其中之一被破坏,它可能会被重新编程为连续传输;这样一个所谓的“笨蛋”,让整辆公交车都无法使用。如果这是动力总成巴士,汽车几乎无法驾驶;驾驶员仍将有一些转向控制,但没有转向或制动的动力辅助。如果汽车高速行驶,则存在严重的事故风险。恶意行为者可能会入侵数百万辆汽车,同时导致数万起道路交通事故,这种可能性是不可接受的,因此必须修补此类漏洞。但是修补是昂贵的。普通汽车可能包含来自 20 家不同供应商的 50-100 个电子控制单元,让它们顺利协同工作所需的集成测试非常昂贵。我将在 27.5.4 节中更详细地讨论这个问题。

攻击不仅限于汽车本身。 2017 年,埃隆·马斯克 (Elon Musk) 对听众说,“原则上,如果有人能够破解所有自动驾驶特斯拉,他们可以说 我的意思是只是一个恶作剧 他们可以说‘把它们都送到罗德岛’美国……那将是特斯拉的终结,罗德岛会有很多愤怒的人。”他的听众大笑起来,三年后才发现他并非完全是在开玩笑。几个月前,一名黑客控制了特斯拉“母舰”服务器,该服务器控制着整个车队;幸运的是,他是一名白帽子,并向特斯拉 [1119] 报告了黑客攻击。在天平的另一端,行为艺术家西蒙·韦克特 (Simon Weckert) 于 2020 年 2 月在柏林周围拉着一辆装有 99 部安卓手机的手推车,无论他走到哪里,谷歌地图都会显示交通堵塞 [1997]。随着高级驾驶辅助系统越来越广泛地依赖云设施,此类间接攻击的范围将会扩大。

外部攻击不需要涉及计算机。如果汽车系统开始为行人和骑自行车的人自动减速,其中一些可能会利用这一点。在印度和南欧的一些地区,行人穿过拥挤的道路,示意汽车停下来,他们也这样做了;看看这种行为是否也出现在伦敦和纽约将会很有趣。

如果可以,公司将利用辅助系统。现在,自动驾驶卡车的最初梦想似乎已经落空,甚至是多辆卡车由一名司机在配送中心之间编队行驶的中间梦想似乎也雄心勃勃,我们是否可以期待通过游说放宽对司机工作时间的法律限制?货运公司可能会争辩说,一旦卡车在高速公路上自动驾驶,司机只需在到达和离开时做真正的工作,所以他应该每班工作十小时而不是八小时。但如果这项技术的最终效果是让卡车司机用同样的钱工作更多的时间,它会

Machine Translated by Google

被怨恨,也许被破坏。

如果 5 级自动化曾经发生过,即使是在受限的环境中 这样我们终于看到了谷歌希望发明的机器人出租车 那么我们将不得不将社交黑客视为安全的一个方面。如果您 12 岁的女儿叫出租车放学回家,那么目前我们有法律形式的保障措施,要求出租车司机对犯罪记录进行背景调查。优步试图避开这些法律,声称自己不是出租车公司,而是一个“平台”;在伦敦,市长不得不禁止他们并在法庭上与他们斗争多年才能让他们遵守。那么 Robotaxis 将如何进行维护工作呢?

也将有责任游戏。目前,汽车公司试图将事故归咎于司机,因此每次事故都变成了哪个司机疏忽大意的问题。但是,如果是计算机在驾驶汽车,那就是产品责任,制造商必须支付费用。围绕辅助驾驶的安全数据存在一些有趣的争论,特别是汽车制造商是否在激活自动驾驶仪的情况下发生车祸,我们将在第 28.4.1 节中讨论。

这么多是完全可以预见的。但是对系统本身的 AI 组件的新攻击呢?例如,您能否通过在桥梁或道路上投射具有欺骗性的图像来迷惑汽车并导致其坠毁?这是很有可能的,而且我已经看到过视觉混乱导致的崩溃。从我的实验室回家的路上,在右转弯处有一所房子,它的主人经常把车停在迎面而来的车道上。晚上,在像英国这样的靠左驾驶的国家,你的驾驶反射是转向对面车辆的左侧,但随后你会发现你正驶向他的花园围墙,于是向右转弯从对面车辆的右侧驶过他的车。最终,一辆大卡车没有及时转向,撞到了墙上。

那么聪明的软件能否以新的方式或更容易让攻击者扩展的方式欺骗机器视觉系统?这就把我们带到了下一个话题,人工智能,或者更准确地说,机器学习。

25.3 人工智能/机器学习

人工智能这个词在不同的时代有不同的含义。
对于像艾伦图灵这样的先驱来说,它的范围从图灵测试到尝试教计算机下棋。到 1960 年代,它意味着文本处理,从 Eliza 到早期的机器翻译,以及 Lisp 编程。在 1980 年代,日本宣布了一项针对“第五代计算”的庞大研究计划,引发了研究热潮,西方国家争先恐后地跟上;大部分努力都投入到基于规则的系统中,Prolog 加入了 Lisp 的行列,成为计算机科学课程中的语言之一。

从 1990 年代开始,重点从具有大量规则的手工系统转变为从示例中学习的系统,现在称为机器学习 (ML)。
早期的机制包括逻辑回归、支持向量机 (SVM) 和贝叶斯分类器;进步是由自然语言处理 (NLP) 和搜索等应用推动的。虽然 NLP 社区开发了自定义方法,但设计支付欺诈检测器或垃圾邮件过滤器的典型方法是收集大量训练数据,编写自定义代码

Machine Translated by Google

提取一些信号,然后根据经验查看哪种类型的分类器对它们最有效。搜索在 2000 年代变得极具对抗性,因为搜索引擎优化公司使用各种技巧来操纵搜索引擎所依赖的信号,而引擎反过来反击,惩罚或禁止使用隐藏文本等不正当技巧的网站。 Bing 是 ML 的早期用户,但谷歌多年来一直避免使用它;从 2000 年到 2016 年退休一直负责搜索的工程师 Amit Singhal 认为,对于一组给定的结果,很难准确地找出众多输入中的哪一个对哪个结果最负责。这使得调试基于机器学习的搜索排名算法变得困难。如果您检测到僵尸网络点击了伊斯坦布尔的餐馆,并想调整算法以将其排除,那么更改一些“if”语句比重新训练分类器 [1300] 更容易。

2011 年开始发生翻天覆地的变化,当时 Dan Cire?san、Ueli Meier、Jonathan Masci 和 Jurgen Schmidhuber 训练了一个深度卷积神经网络,在识别手写数字和汉字方面表现与人类一样好,并且在交通标志方面优于人类 [435] .次年,Alex Krizhevsky、Ilya Sutskever 和 Geo Hinton 使用类似的深度神经网络 (DNN) 在对 120 万张图像进行分类时获得了破纪录的结果 [1098]。比赛已经开始,其他研究人员蜂拥而至,“深度学习”开始在各种任务中受到广泛关注。最引人注目的结果出现在 2016 年,当时 David Silver 和谷歌 Deepmind 的同事制作了 AlphaGo,击败了世界围棋冠军李世石 [1737]。这引起了世界的关注。

在此之前,很少有研究生想学习机器学习;从那以后,很少有人想学习其他任何东西。本科生甚至在课堂上关注概率和统计,这在以前被认为是一件苦差事。

25.3.1 机器学习与安全

机器学习和安全之间的互动可以追溯到 1990 年代中期。正如我在第 21.3.5 节中描述的那样,恶意软件编写者开始使用诸如多态性之类的技巧来逃避反病毒软件中的分类器;正如我在第 12.5.4 节中所述,银行和信用卡公司开始使用机器学习来检测支付欺诈;正如我在 22.2 节中提到的,电话公司也将其用于第一代手机。随着 Internet 在 20 世纪 90 年代中期向公众开放,垃圾邮件的到来为垃圾邮件过滤器创造了市场。对于大型邮件服务提供商而言,手工制定的规则的扩展性不够好,尤其是在僵尸网络出现并且垃圾邮件成为电子邮件的主体之后,因此垃圾邮件过滤成为了一个重要的应用程序。

Alice Hutchings、Sergo Pastrana 和 Richard Clayton 调查了机器学习在此类系统中的使用,以及坏人想出的欺骗他们的技巧 [939]。由于垃圾邮件过滤将用户反馈作为其基本事实,垃圾邮件发送者学会了将垃圾邮件发送到他们在大型网络邮件公司控制的帐户,并将其标记为“不是垃圾邮件”;现在使用其他统计分析机制来检测这一点。使分类器的训练数据中毒是一种非常普遍的攻击。另一种是寻找价值链中的薄弱点:机票欺诈者购买一张无害的机票,通过欺诈检查,然后在出发前将其更改为前往高风险目的地的机票。并且在地下论坛上对此类技术进行了热烈的讨论

Machine Translated by Google

不良行为者交易的不仅仅是服务,还有吹嘘和小费。 Battista Biggio 和 Fabio Rolli 提供了更多的技术背景:2004 年,垃圾邮件发送者发现他们可以通过改变某些词来混淆垃圾邮件过滤器中的早期线性分类器,并且军备竞赛从那里开始 [241]。

事实证明,这些攻击思想可以推广到其他系统,并且有
其他攻击也是如此。

25.3.2 对 ML 系统的攻击

机器学习系统至少有四种类型的攻击。

首先,你可以毒化训练数据。如果模型在使用中继续自我训练,那么很容易让它误入歧途。 Tay 是微软于 2016 年 3 月在 Twitter 上发布的一款聊天机器人;巨魔立即开始教它使用种族主义和攻击性语言,仅 16 小时后它就被关闭了。

其次,您可以在推理阶段攻击模型的完整性,例如使其给出错误答案。 2013 年,Christian Szegedy 及其同事发现,在 2012 年被发现可以很好地对图像进行分类的深度神经网络很容易受到对抗性样本的攻击 图像稍微被扰乱就会被严重错误分类 [1857]。这个想法是选择一个扰动来最大化模型的预测误差。事实证明,神经网络有很多这样的盲点,它们以不明显的方式与训练数据相关。决策空间是高维的,这使得盲点在数学上不可避免 [1706];并且对于神经网络,决策边界是复杂的,使它们不明显。研究人员很快想出了真实世界的对抗性例子,从会导致汽车视觉系统将 30 英里/小时的速度标志误读为 60 英里/小时的小贴纸,到会导致戴眼镜的男性被误认为女性的彩色眼镜,或根本不被认可[1720]。在恶意软件检测领域,人们发现非线性分类器(如 SVM 和深度神经网络)实际上并不比线性分类器更难规避,前提是你做得对 [241]。

第三,Florian Tram`er 和他的同事表明,您可以在推理阶段攻击模型的机密性,方法是让它对多个探测输入进行分类并构建一个连续更好的近似值。结果通常是目标模型的良好工作模仿。与制造真实商品一样,仿制品通常更便宜;从头开始训练大模型可能会花费很多。这种近似攻击不仅适用于神经网络,还适用于其他分类器,例如逻辑回归和决策树 [1901]。

更重要的是,许多攻击被证明是可转移的,因此攻击者不需要完全访问模型(所谓的白盒攻击)[1900]。许多攻击可以在一个模型上开发,然后针对另一个在相同数据甚至类似数据上训练过的模型发起(黑盒攻击)。

盲点是训练数据的函数,所以为了使攻击不易转移,你必须付出努力。例如,Ilia Shumailov、Yiren Zhao、Robert Mullins 和我曾尝试在神经网络中插入密钥,使盲点出现在不同的地方,具有不同密钥的模型容易受到不同的对抗样本的影响 [1733]. Kerckhos

Machine Translated by Google

原则适用于机器学习,就像安全领域的几乎所有其他地方一样。

机密性攻击的一个变体是提取敏感的训练数据。
大型神经网络包含大量状态,处理异常值最简单的方法通常只是记住它们。因此,如果某些企业声称基于百万医疗记录训练的分类器不是个人数据,因为它是“统计机器学习”,请小心。我们在第 11.3 节中讨论的将机器学习与差异隐私相结合的方法是一个活跃的研究主题 [1493]。

最后,您可以拒绝服务,一种方法是选择会导致分类器花费尽可能长的样本。 Ilia Shumailov 及其同事发现,人们通常可以通过向分类器提出难题来拒绝服务。

给定一个直通管道,就像在一个典型的图像处理任务中一样,一张令人困惑的图像可能会多花费 20% 的时间,但在更复杂的任务中,例如自然语言处理,您可以调用异常处理并使速度减慢数百倍 [1730]。

更复杂的攻击跨越了这些类别。例如,在线广告商和广告拦截软件供应商之间存在军备竞赛,随着广告商采用越来越复杂的网页渲染方式来迷惑拦截者,拦截者开始在渲染的页面上使用图像处理技术页面来发现广告。然而,这使它们对使用对抗性样本的广告商开放,要么逃避过滤器,要么导致它错误地阻止页面的另一部分[1899]。

那么如何在现实世界中安全地使用机器学习呢?这是我们仍在学习的一些事情,但有些事情我们可以说。首先,必须采用系统安全方法并端到端地查看问题。正如我们对 Web 服务的输入进行清理、进行渗透测试并拥有负责任的披露和更新机制一样,我们需要对 ML 系统做同样的事情 [659]。

其次,我们需要借鉴过去二十年在信用卡欺诈、垃圾邮件和入侵检测等主题上的工作经验。正如我们在 21.4.2.2 节中提到的,ML 系统在现实世界的网络入侵检测中基本上是无效的； Robin Sommer 和 Vern Paxson 是第一个给出很好解释的人。他们讨论了训练数据的缺乏、理论与实践之间的距离、评估的困难、错误的高成本以及最重要的是无法应对新的攻击 [1802]。将有能力的对手排除在复杂的企业网络之外并不是人工智能一直擅长的问题。

不过,重点可能偶尔会发生变化。如果我们想降低新的对抗性攻击造成实际损害的可能性,我们可以根据具体情况采取多种措施。一种是简单地使分类器失谐;这是至少一种用于汽车的机器视觉系统的方法。通过降低它的敏感度,你可以让它更不容易被欺骗,然后你可以用雷达和超声波等其他传感器来补充它,这样视觉系统本身就不那么重要了。另一种方法是朝另一个方向前进,通过使系统的 ML 组件足够脆弱,以至于其他组件可以检测到攻击 于是您切换到防御性操作模式,例如低灵敏度跛行回家模式或

Machine Translated by Google

停下来等人开车。换句话说,你开始建立情境意识。这就是我们在现实生活中的行为方式;正如我在第 3.2.5.1 节中讨论的那样,祖先的进化环境教会我们在感觉到敌对意图和违反部落禁忌等触发因素时要格外小心。因此,我们尝试使用经过训练的神经网络,以便将许多输出和激活视为禁忌和避免;如果这些禁忌中的任何一个被打破,就可以怀疑是攻击 [1733]。

根本问题在于,一旦我们开始让机器学习模糊代码和数据之间的界限,并且系统变成数据驱动的,人们就会去玩弄它们。这给我们带来了机器学习与社会互动的棘手问题。

25.3.3 机器学习与社会

自 2016 年以来,人们对机器学习的兴趣激增,并在大众媒体中将其描述为“人工智能”,引发了很多关于伦理的猜测。例如,哲学家丹·丹尼特 (Dan Dennett) 以道德为由反对不朽、有智慧但没有意识的人的存在。

但公司已经符合这个定义!公司不当行为的历史表明,公司的行为确实可能非常糟糕(我们在 12.2.6 节中讨论了一些例子)。最强大的机器学习系统属于谷歌、亚马逊、微软和 IBM 等公司,它们都曾与权威发生过争执。 ML、大数据和垄断之间的相互作用增加了政府在考虑如何监管技术时需要解决的问题。一方面是科技专业的 ML 产品现在正在成为自己的平台,并被许多初创公司用来解决特定的现实世界问题 [658]。

一个贯穿各领域的问题是偏见。普林斯顿的土耳其研究生 Aylin Caliskan 注意到,从土耳其语到英语的机器翻译带有性别偏见;尽管土耳其语没有语法性别,但土耳其语句子的英文翻译会将医生指定为“他”,将护士指定为“她”。在进一步调查中,她和她的主管乔安娜布赖森 (Joanna Bryson) 和阿尔温德纳拉亚南 (Arvind Narayanan) 发现,基本上所有使用的机器翻译系统不仅存在性别歧视,而且还存在种族歧视和恐同 [369]。事实上,大量基于机器学习的自然语言系统吸收了它们训练数据的偏见。如果大平台的 ML 引擎可以通过数百家下游公司所依赖的系统来阻止偏见,那么肯定存在公共政策问题。

一个相关的政策问题是红线。保险公司在使用邮政编码级别的理赔统计数据来确定保费水平时,发现许多少数民族地区的保费高昂或被排除在外,违反了反歧视法。我在 2008 年的第二版中写道:“如果你基于数据挖掘技术构建入侵检测系统,你将面临严重的歧视风险。如果你使用神经网络技术,你将无法向法庭解释你的决定背后的规则是什么,因此为自己辩护可能会很困难。不透明的规则还可能违反欧洲数据保护法,该法赋予公民了解用于处理其个人数据的算法的权利。”

Machine Translated by Google

第二个交叉问题是蛇油,AI/ML 淘金热催生了数以千计的初创公司,其中许多在营销方面比在产品方面更强大。 Manish Raghavan 及其同事对用于就业筛选和招聘的“AI”系统进行了调查,发现数十家公司声称他们的系统可以将新员工与公司的要求相匹配。大多数人声称他们不歧视,但由于很少有雇主保留有关员工绩效的全面且可访问的数据,因此完全不清楚如何培训此类系统,更不用说使用此类系统的公司如何为歧视诉讼辩护 [1571]。

申请者很快就会学会玩弄这个系统,比如将“牛津”或“剑桥”这个词用白字写进他们的简历中。谨慎的雇主会要求更透明的机制,并设计独立的指标来验证其结果。即使这样也很重要,因为机器学习可以发现我们不理解的相关性。

Arvind Narayanan 对 AI 中的蛇油进行了有趣的分析 [1382]。 “AI”甚至“ML”都是一整套技术的通用术语。其中一些已经取得了真正的进步,比如用于人脸识别的 DNN,甚至是 AlphaGo。因此,公司利用这种炒作,在他们销售的任何产品上贴上“AI”标签,即使其机制使用的是一个世纪前的统计技术。深入挖掘,Arvind 认为机器学习系统可以分为三类:

1. ML 在感知任务上取得了真正的进步,例如人脸识别(见第 17.3 节)、Shazam 等产品对歌曲的识别(见第 24.4.3 节)、扫描医学诊断和语音到文本 在所有这些方面,它都获得了熟练人员的能力;

2. ML在内容推荐、垃圾邮件和仇恨言论识别等判断任务上取得了一定进展。这些有许多硬边案例,即使是熟练的人也不同意。执行它们的系统通常依赖于大量的人工输入 从点击“报告垃圾邮件”按钮的十亿电子邮件用户到大型科技公司雇用的数万名内容审核员;

3. ML 在社会预测任务上没有取得进展,例如预测员工绩效、学校成绩和未来犯罪行为。

   Matthew Sagalnik 和 400 多名合作者进行的一项非常广泛的研究得出的结论是,只要可以预测生活结果,也可以使用基于少数变量的简单线性回归来完成 [1638]。

这是一个可证伪的说法,所以我们会随着时间的推移看看它的准确性,如果这本书在 2030 年有第四版,我们就会有更多的数据。与此同时,研究的一个主要主题将是寻找更好的人机协作方式。直觉上,我们希望人们从事涉及判断力的工作,而机器从事无聊的研究 ;但让它真正起作用可能比看起来更难。人们通常最终成为机器的仆人,据一家风险投资公司称,40% 的“人工智能”初创公司实际上并未以任何实质性方式使用机器学习;他们只是乘着炒作的浪潮雇用幕后人员 [1960]。不管怎样,路上会有很多颠簸,也会有很多关于道德和政治的辩论。

Machine Translated by Google

也许接近道德的最好方法就是这样。现在在 AI 伦理背景下讨论的许多问题都是多年前针对使用传统统计方法对个人信息数据库进行的研究而产生的。

（事实上，线性回归已经连续使用了大约一个世纪；它们刚刚被重新命名为机器学习。）因此，我们的第一个停靠点应该是现有的法律和政策。当我们在第 10.4.5.1 节中讨论基于记录的健康和社会政策研究背景下的伦理时，我们观察到许多问题的出现是因为 IT 公司及其客户忽视了医生、教师和其他人多年来积累的智慧处理纸质记录。同样的错误现在正在重演，并且像以前一样以围绕“创新”和“颠覆”的销售炒作为借口。

在预测哪些儿童可能会犯罪的情况下，多年来人们都知道此类指标可能会造成严重的污名化。在第 10.4.6 节中，我们注意到，如果您告诉老师哪些孩子接触过社会服务，那么老师对他们的期望就会降低。儿童福利和隐私法都反对共享此类指标。如果无知的管理员购买了声称使用使 AlphaGo 击败李世石的不可思议的魔法进行预测的软件，那么危害会更大吗？至于“预测性警务”，研究表明，这可能只是让计算机证明“围捕通常的嫌疑人”政策合理的另一种方式 [677]。 （在第 14.4 节中，我们讨论了宵禁标签如何产生这种效果。）在保释听证会上以及量刑听证会上，使用 ML 技术向法官提供有关嫌疑人是否构成逃跑风险或再判风险的建议时，也会出现类似的问题关于嫌疑人是否危险。此类技术可能会传播现有的社会偏见和权力结构，并为立法者提供继续实施无效但民粹主义政策的借口，而不是推动他们解决根本问题。

尽管如此，ML 还是有可能打破多年来在监视、隐私和审查制度等问题上出现的一些平衡，因为它为已经强大的参与者提供了更强大的工具，并为旧的滥用创造了新的借口。许多国家已经限制闭路电视摄像机的使用；既然人脸识别系统可以识别行人，我们是否需要对他们进行更多限制？正如我们在第 17.3 节中看到的，许多城市（包括旧金山）已经决定答案是肯定的。

在第 11.2.5 节中，我们讨论了位置和社交数据现在如何使匿名变得非常困难，以及人们的 Facebook 数据如何被挖掘用于政治广告定位。 ML 技术通过发现通信模式 [1719]，使进行 trac 分析变得更加容易；事实上，警察和情报机构越来越依赖于第 21.7、23.3.1 和 26.2.2 节中讨论的那种 trac 和社交网络分析。

简而言之，针对机器学习的指控是，它是一种技术，有助于巩固科技专业人士的权力，同时将隐私和监视之间的平衡推向监视，并以其他方式促进专制政府。谷歌和微软正在资助大型研究项目以开发人工智能造福于社会，这或许说明了这一点。

那么，在我们现在生活的这个电子村里，我们实际上可以做些什么来获得一些隐私呢？

Machine Translated by Google

25.4 PETS 和操作安全

即使您没有在 Facebook 上脱口而出,社会结构（谁和谁一起出去玩)也能说明很多,而且变得更加明显。在第 11.2.5 节中,我们讨论了一项研究,该研究表明,只要 4 个 Facebook 点赞,细心的观察者就能在大部分时间判断出你是异性恋还是同性恋,以及这种观察如何导致剑桥分析丑闻等事件,选民的偏好被秘密和详细地记录下来。

即使您根本不使用 Facebook,关于谁联系了谁的 trac 数据也会向有权访问它的人泄露很多信息,正如我们在第 11.4.1 节中讨论的那样。
这可能会给与权威发生冲突的人（例如举报人)带来麻烦。匿名在这里有时是一个有用的工具。学术权威的滥用通过对教授的匿名学生反馈和对会议论文提交的匿名审阅来抵消。如果你的雇主支付你的健康保险,你可能想用现金购买 HIV 检测试剂盒并在网上匿名获得结果,因为你接受检测这一事实本身就说明了一些事情,即使结果是否定的。隐私也可能是言论自由的必要前提。试图在政治或宗教领域进行创新的人可能需要在公开之前发展他们的学说并建立他们的人数。然后还有反对派政客为当时的政府挖空心陷阱,他们的担忧更具战术性。

此类活动对开放社会的重要性如此之高,以至于我们将隐私和言论自由视为相互关联的人权。我们还制定法律来保护举报人。但这在实践中如何实现呢?

在前技术社会中,两个人可以与其他人走一小段距离并进行对话,而不会留下任何确凿的证据证明所说的话。如果爱丽丝声称鲍勃批评了国王,那么鲍勃总是可以反其道而行之 爱丽丝提议举行示威以增加议会的权力,而他出于忠诚而拒绝了。

换句话说,许多通信都是可以否认的。似是而非的否认仍然是当今某些通信的重要特征,从日常生活到最高级别的情报和外交。它有时可以通过约定来固定:例如,英国的诉讼当事人可以写一封标有“不影响”的信件给另一名提出和解的信件,并且这封信件不能用作证据。但大多数情况下都缺乏这样明确和方便的规则,而且通信的电子性质通常意味着“只是走出去一分钟”不是一种选择。然后怎样呢?

一个相关的问题是匿名性。在工业革命之前,大多数人都住在小村庄里,搬进城镇是一种解脱 实际上是一场革命。你可以改变你的宗教信仰,或者投票给土地改革候选人,而你的房东不会把你赶出 你的农场。在很多方面,互联网的影响已经把我们带回了一个“电子村”:电子通信不仅缩短了距离,而且在某些方面也缩短了我们的自由。

技术可以提供帮助吗?为了使事情更具体一些,让我们考虑一些有特定隐私问题的人。

Machine Translated by Google

1. 安德鲁是得克萨斯州的一名传教士,他的网站在伊朗吸引了许多皈依者。那个国家处决改变宗教信仰的穆斯林公民。他怀疑与他联系的一些人并不是真正的皈依者,而是追捕叛教者的宗教警察。他无法区分警察和真正的皈依者。他应该使用什么样的技术与皈依者私下交流?

2. 贝拉是您 10 岁的女儿,她的老师警告她在网上保持匿名。你应该给她什么样的训练?

3. Charles 是一位心理分析师,他看到私人患者患有抑郁症、焦虑症和其他问题。以前他在镇上一所不起眼的房子里行医,他的病人可以谨慎地参观。自封锁以来,他不得不使用 Skype 和 Zoom 等工具。保护患者隐私的谨慎做法是什么?

4. 戴是越南的一名人权工作者,与试图建立独立工会、小额信贷合作社等的人接触。警察经常骚扰她。她该如何与同事沟通?

5. Elizabeth 是一家投资银行的分析师,为合并提供建议。她想要调查收购目标的方法,而不会让目标听到她的兴趣 甚至不知道任何人都感兴趣。她的对手是其他公司里和她一样的人。

6. Firoz 是一名住在德黑兰的男同性恋者,在德黑兰,同性恋是死罪。
他想要某种方式来下载色情内容,或许还可以在不被绞死的情况下联系其他男同性恋者。

7. Graziano 是巴勒莫的一名地方法官,他设立了一条热线电话,让人们向当局举报有关黑手党活动的消息。他知道未来一些驻扎在办公室的警察将得到黑手党的报酬 而且潜在的线人也知道这一点。他如何限制腐败警察可能造成的损害?

8. Hristo 帮助难民进入英国,以便他们可以申请庇护。他的大多数客户都在逃离中东和北非的战争或糟糕的政府。他在比利时开展业务,根据天气情况将客户送上卡车或快艇。他需要与法国、英国和其他地方的同事进行协调。尽管受到各种安全和情报机构的监视,他们如何做到这一点?

9. 艾琳是一家好斗报纸的调查记者,她邀请举报人与她联系。她梦想登陆下一个 Ed Snow 巢穴。她应该做些什么准备,以防她确实被政府极力试图揭露的主要消息来源联系?

10. 贾斯汀正在竞选民选总统。艾琳很乐意挖他家人的污点。还有许多其他人想阅读他的电子邮件,从他的社交媒体帐户发送种族主义推文,或将他的竞选战利品汇给朝鲜。他怎么能挫败他们呢?

Machine Translated by Google

隐私不仅仅是加密消息。如果安德鲁告诉他的皈依者下载并使用 Wickr，那么伪装成皈依者的警察间谍将获得国家防火墙以检测任何使用它的人。 Andrew 必须让他的 trac 看起来无伤大雅 这样即使警察知道叛教者的 trac 是什么样子，也无法发现皈依者。如果只有几十个人使用 Wickr，警察就可以破门而入。因此，这不仅仅是关于 Wickr 是否比 Signal 或 Skype 更安全的产品，而是该国有多少人使用它。

虽然技术措施可以解决安德鲁的部分问题，但它们对贝拉的问题用处不大。对孩子来说，一个风险是他们会说一些粗心的话，这可能会让他们以后感到尴尬。另一个是关于儿童安全的政治和媒体恐吓妨碍了他们的福利。你的大部分努力将用于教育她。随着贝拉的成长，她将不得不熟练使用其他同龄人使用的工具；很快她就会更多地从他们那里采纳她的安全程序，而不是从你那里采纳。你必须传授理解，而不是仪式。

攻击的强度会有很大差异。 Andrew 和 Firoz 可能只会面临零星的兴趣，而 Graziano、Hristo 和 Justin 都有有能力的积极对手。至于戴，她经常受到监视。她使用匿名通信不是为了保护自己，而是为了保护尚未引起警方注意的其他人。

存在截然不同的激励机制。 Andrew、Charles、Dai、Graziano 和 Irene 不厌其烦地保护他们打交道的弱势群体，而 Firoz 感兴趣的网站并不关心他的安全。 Andrew、Dai、Graziano 和 Hristo 都必须考虑不诚实的内部人员。在贾斯汀的案例中，这是粗心的内部人士：俄罗斯人想给艾琳的有趣故事存在于他的竞选志愿者的个人账户中，以及难以包含在任何有组织的活动中的朋友和家人的个人账户中防守努力。

成功和失败有不同的门槛。只有在警方排除合理怀疑证明对他不利的情况下，Hristo 才能被判入狱；如果艾琳能在证据平衡的基础上为诽谤诉讼辩护，她就可以打败贾斯汀；而仅仅是怀疑对伊丽莎白或菲罗兹来说可能是个坏消息。失败的代价也不同 ：如果伊丽莎白搞砸了，她可能会损失一些钱，而贾斯汀可能会失去他的职业生涯，菲罗兹可能会失去生命。

我们在第 22.2.1 节中讨论了那些不想让他们的电话被跟踪的人如何购买预付费手机，使用一段时间，然后扔掉。
但是这些燃烧器，正如它们有时被称为的那样，很难正确使用；甚至基地组织也做不到。那么在线硬隐私的现状如何呢？

25.4.1 匿名消息设备

正如我们在 2.2.1.10 节中讨论的那样，调查人员通常从 trac 分析中获得大部分信息。不管人们是使用电子邮件、消息服务还是普通的老式电话服务，访问社交图谱都可以让警察绘制友谊网络 营销人员在他们需要时也会这样做

Machine Translated by Google

可以得到他们的手[598]。在过去,加密您的电子邮件轨迹可能很危险;如果您是该国仅有的 20 个使用 PGP 的人之一,那么您就会成为嫌疑人。现在大多数人都使用默认使用 TLS 加密的网络邮件服务,情况变得更加复杂,但适用相同的原则。

像 Hristo 这样受到政府监控的人了解到,像 WhatsApp 或 Signal 这样的普通隐私应用程序本身是不够的,即使很多其他人出于无害目的使用它们也是如此。假设 Hristo 使用 Signal 安排 Kevan 带着八个人乘坐快艇横渡英吉利海峡。但是,如果一名皇家海军快艇逮捕了凯文,并且他们在他的手机上发现了赫里斯托的信息,他将面临引渡。如果 Kevan 或 Hristo 也使用手机与家人聊天,这可能有助于警方使用 trac 分析绘制他们的网络图。不仅仅是让网络难以追踪的问题,还有当人们被抓到时可以扣押哪些证据的问题。戴和格拉齐亚诺的卧底特工也面临着类似的问题。

因此,我们看到了“加密电话”市场的发展,它不仅提供加密消息,而且还尝试支持操作安全 (Opsec)。我们在 3.3.4 节中讨论了企业环境中的 Opsec,但它在这里更重要。第一款公开发售的加密电话可能是 Silent Circle 于 2014 年推出的 Blackphone,它被出售给政府机构、特种部队和人权工作者。此后出现了许多相互竞争的系统。 Ed Caesar 描述了一些从德国的 Cyberbunker 推广加密电话业务的人,在 2019 年 9 月遭到突袭并关闭之前,Cyberbunker 是该国最大的非法网站托管者 [364]。手机通常经过修改,因此您无法运行应用程序（可能会监视您）;他们可能禁用了麦克风和摄像头,因此 GCHQ 无法将它们变成监控设备； GPS 也可能被禁用;它们不能被标准的警察取证亭读出;它们是一个封闭系统的一部分,该系统由电话和消息服务器组成,在该系统中,您不是通过电话号码而是通过用户 ID 来识别另一方。加密电话公司发现,有些人愿意为一部手机支付超过 1000 美元或欧元的费用,而对于相关服务的六个月订阅费用也是如此。市场包括各种各样的人,从加密货币运营商和间谍到洗钱者再到毒贩。网络效应也适用于隐蔽社区； Hristo、Kevan 和该团伙的其他成员都需要使用相同的系统。由于一些走私者也走私毒品,而一些走私者赚到的钱足以聘请同样为加密货币人群工作的高级税务会计师,网络效应可能会吸引各种各样的人,他们出于好的和坏的原因寻求隐私免受国家监视坏的。

新出现的模式是,由于网络效应,一个加密电话系统得到了越来越广泛的使用,直到足够多的用户成为警察目标并且当局将其摧毁。为了非英国读者的利益,我可能会在这里提到左派和右派报纸对赫里斯托和他的人类货物的看法有所不同 。虽然一些移民社区将 Hristo 的行动视为家庭团聚服务,但保守派媒体对难民进行污名化,部长们已将移民犯罪作为这些机构的优先事项,而不是有组织的贪得无厌的犯罪。那么 Hristo 可以买什么来让 GCHQ 支持他呢?

直到 2016 年,市场领导者是荷兰公司 Ennetcom,它使用

Machine Translated by Google

加拿大的消息服务器专用网络,支持匿名用户 ID。同年 4 月,荷兰和加拿大当局突击搜查并逮捕了与 CyberBunker 有牵连的所有者。 2017年,轮到PGP Safe;四名荷兰人被捕 [1081]。次年,荷兰警方还声称破解了一个名为 Iron Chat [792] 的加密电话系统。 2018 年,市场领导者是一家名为 Phantom Secure 的公司;美国、澳大利亚和加拿大当局关闭了该系统 [1133]。

其首席执行官文森特拉莫斯承认向全球毒贩提供手机,在他的量刑听证会上,检察官宣读了他发给同事的一条信息:“我们他妈的是有钱人……得到他妈的路虎揽胜品牌新的。

因为我刚刚关闭了很多业务。这周的人。锡那罗亚卡特尔,就是这样”[278]。他入狱九年。下一个市场领导者 EncroChat 使用经过修改的 Android 手机。 2020 年,法国和荷兰警方入侵了其主服务器,并用执法恶意软件感染了全球所有使用的 50,000 台设备,这些恶意软件将他们的信息实时复制给了警方。 6 月 13 日,EncroChat 意识到他们被黑了,并建议他们的客户立即扔掉他们的手机 [1922]。整个欧洲都有数百人被捕 [572]。

所以像格拉齐亚诺这样的警察有一个标准的剧本来摧毁加密电话系统。但他也可以使用它们来保护那些仍然留在帮派和他们游弋的社区中的消息来源。

事实上,当 PGP 在 1990 年代首次出现时,临时 IRA 在他们反对英国在北爱尔兰统治的叛乱中采用了它。直到那时,让警察头疼的一个大问题一直是与爱尔兰共和军线人进行不引人注目的定期接触,这些线人生活在一个仇恨警察的民族主义社区,线人在那里被杀。 PGP 使联系变得容易。告密者只需将他的私钥告诉他的上司,警察就可以收集他的所有踪迹。他甚至可以通过向自己发送一封加密的电子邮件来报告。

25.4.2 社会支持

记者艾琳的任务可能最艰巨。如果一位高级公务员找到她,想揭露政府最近的愚蠢行为,那么一旦故事一出现,“内奸”就会开始。她的线人 让我们称她为丽兹吧 现在将被警察和情报机构追捕。

Irene 如何帮助 Liz 将被识别、解雇和起诉的可能性降到最低?我们在第 2.3.6 节中简要讨论了举报,其中我们看到技术安全通常只是举报者面临的问题之一 而且往往不是最严重的。

最大的问题是建立信任,这是一个双方面的过程。 Irene 需要将 Liz 作为消息来源进行评估。她有真实的故事要讲吗?她为什么要说?这是一个半授权的泄密,她在她的部长的默许下作为政治游戏的一部分提供?如果有人以诽谤罪起诉她,她的故事能否站得住脚?这是一种挑衅,旨在诋毁艾琳或她工作的报纸吗? Liz 是否脆弱,需要情感支持?当故事出来时,还有谁会泄露它?如果一百个人都可以泄露它,你可以谈论匿名;如果匿名集的大小只有 10,那么你更多的是在谈论似是而非的推诿,Irene 会想和 Liz 谈谈会发生什么

Machine Translated by Google

当 PM 的手下审问她时。但很多情况下,一旦事情真相大白,举报人就会彻底暴露。例如,如果 Liz 的投诉是一位部长试图强奸她,那么与 Irene 的对话将是关于获得支持以及人们是否会相信她,而不是关于如何使用 Signal。

因此,最佳做法是在 Liz 联系后,Irene 尽快亲自与 Liz 会面。如果 Liz 可能成为国家行为者的目标,但有合理的机会保持匿名,Irene 可以给她一个一次性电话以建立独立于她现有的家庭和工作设备的联系链。如果 Liz 是故事曝光后的十名嫌疑人之一,并且总理开始对安全局局长大喊大叫,那么她最好假设这十个人的所有已知设备都会在下午茶时间受到损害。

当埃德·斯诺登决定揭发非法监控时,他最初很难让记者使用 PGP 加密。之后,许多报纸争先恐后地为举报人提供联系他们的技术手段,公布了 PGP 密钥、使用 Signal 的记者的手机号码,以及一个名为 SecureDrop 的设施,使人们能够上传文件。

Mansoor Ahmed-Rengers、Darija Halatova、Ilia Shumailov 和我对此类机制进行了研究,发现它们存在两种类型的问题 [31]。首先,这种机制很难使用。我们在 3.2.1 节中讨论了安全可用性研究是如何从使用 PGP 的困难开始的,问题仍然存在。其次,举报人需要了解危害,以便设计合理的操作安全程序,但典型的报纸不会像本章那样讨论它们。因此,Irene 可能不仅想给 Liz 一个一次性电话,还想给她一个关于如何使用 Tails 和 Tor 等工具将文件上传到 SecureDrop2 的培训课程。 （加密电话会更有用,但 Irene 可能没有预算,如果 Liz 被抓到一部,它可能是赠品。）

但是,将 Liz 视为 Irene 的典型消息来源是错误的。大多数告密者都属于一号匿名组,他们披露的不是国家机密,而是欺诈和滥用职权。在第 12.2.6 节中,我们看到举报人比审计员或监管机构阻止了更多真正严重的欺诈行为。但揭露错误行为的决定通常可能会带来一些个人成本,例如被解雇或被污名化。社会支持往往是关键。

只有在几名被哈维温斯坦强奸的女性鼓起勇气说出来之后,其他数十名女性才站出来。

支持对于我们的许多其他用户也至关重要。心理分析师查尔斯知道,他可以为他的病人（我们可以称之为玛丽)提供隐私,这对治疗工作至关重要。从办公室到视频会议的转变不仅会带来一些（小的)实际风险,而且会使双方都难以理解隐私,从而削弱其在治疗中作为促进者的作用。玛丽可能担心,如果她的雇主发现她正在接受治疗,她可能会被同事污名化或错过晋升机会。在大多数情况下

2即便如此,记者仍应注意更多信息,例如现代打印机嵌入文档中的机器识别码,我们在第 24.4.3 节中对此进行了讨论。他们被用来追踪 Reality Winner,一名 NSA 告密者泄露了一份描述俄罗斯干涉 2016 年大选的 NSA 文件,并被判入狱 63 个月 [174]。

Machine Translated by Google

恐惧会比实际风险大得多,但有时风险可能是真实的:她可能是 Dai、Irene 或 Liz。因此,治疗环境必须让她平静下来并激发信心。 Charles 无法通过 Irene 在他们第一次见面时可能给 Liz 的那种详细的 opsec 简报来开始 的关系。隐私建议(如果有的话)可能必须与其他支持一起点滴提供。

与贝拉这样的孩子打交道时,首要任务还在于为他们提供一个可以学习和发展的平静和安心的环境。明智的父母会识破围绕儿童安全的危言耸听;儿童被陌生人绑架和杀害的比率约为每年千万分之一,这一比率如此之低,以至于理性的人会忽略它。

作为父母,您的使命是帮助您的孩子成长为有能力的公民,而不是训练他们畏惧想象中的怪物。

维权人士戴也是她试图招募的人的支持者。她的案子比查尔斯的要棘手得多,因为当局正试图阻止她发挥作用。我猜她为当局所知,并受到间歇性监视。

像 Dai 这样的人权工作者确实使用 Skype、Tails、Tor 和 PGP 等常用工具来保护他们的流量,但他们所遭受的攻击不仅仅是技术上的;它们是间谍小说的原型。警察偷偷进入他们的家,植入可以窃取密码的 rootkit,并植入房间窃听器来窃听对话。当他们对电话进行加密时,他们不得不怀疑秘密警察是否从一个隐藏的麦克风中获取了通话的一面(或两面)。有时麦克风并没有那么隐蔽;我们从警察积极分子那里听说,他们公然站在他们的房子外面,用猎枪式麦克风指着窗户。

反击此类攻击需要 tradecraft 反过来。其中一些就像在间谍电影中一样:留下告密者以检测秘密进入,始终随身携带笔记本电脑,并在难以窃听的地方进行敏感对话。它的其他方面不同:人权工作者(像记者但不像间谍)需要避免违法,他们还需要培养多种支持结构 不仅仅是为下游的新兵提供秘密支持,而且接受公开的支持海外非政府组织和政府的支持。

为了招募新兵,他们还需要 在间歇性观察下 与自己没有受到怀疑的人进行秘密接触。 Dai 的情况与 Charles 的情况相反,因为当她招募一名新员工时,对他们进行贸易技能培训是入职、社会化和支持过程的一部分。

如果你想了解贸易工艺中哪些有效,哪些无效,那么人权工作者就是可以与之交谈的人。 (间谍和走私者可能知道更多,但他们不说。) 新出现的画面是,警察和非暴力政府反对者的行为都植根于社会的运作方式,并随着时间的推移而演变。这是一个复杂的游戏,除了最极权主义的统治者之外,所有人都试图边缘化、驯服或拉拢他们的对手,而反对派运动则做出回应。任何开始对不受欢迎的统治者过于友善的运动都会失去信誉并被其他人取代。拥有最好的 opsec 的团体将能够增长最快,而最激进的团体可能具有最大的信誉。逼得太紧,

Machine Translated by Google

非暴力反对可能会引发公开叛乱或暴力恐怖主义（谴责非暴力反对为“恐怖主义”的统治者可能会招致这种情况）。所以一个聪明的秘密警察头子会放过戴，看她干什么，打一场持久战；普京的哲学是容忍叛乱运动，直到你弄清楚如何领导他们。以防万一以后事情升温，他会节省使用他的一些能力，所以他有一些她还没有制定对策的储备 stu 。

25.4.3 在土地上生活

Irene、Charles 和 Dai 可能会发现他们的隐私策略受到他们必须给予或接受的支持类型的影响，但他们还有其他共同点 他们必须最明智地利用可用的资源，而不是购买或购买构建专用工具。我们或许可以称其为生活在土地上3。

在过去，隐蔽性可能意味着隐藏在众目睽睽之下。每个国家的精英都有出没的地方，所以如果一个高级公务员想见一个著名的记者，他们可以在伦敦的绅士俱乐部或弗吉尼亚的乡村俱乐部敞开心扉地聊天，而不会引起任何人的注意。这种机制允许人们谨慎地联系，同时建立信任。

因此，当尝试即兴进行匿名交流时，首先要问的是您已经共享了哪些俱乐部或平台。中国是一个棘手的案例，它阻止了我们在防火墙上熟悉的大部分服务。

即使在那里，我们也发现平台对具有加密通信的用户内容开放：两个例子是 Linkedin 和亚马逊书评。就伊朗而言，安德鲁将不得不弄清楚 Skype 和 Signal 等消息系统是否在那里得到了足够广泛的使用，以免引起怀疑。

您必须考虑的第二件事是威胁模型。我们许多用户的一个共同点是间歇性威胁：大多数时候根本没有威胁，但偶尔可能会变得严重。即使是庞大的秘密警察部队也只能同时处理这么多文件。大多数时候，没有人对 Mary 或 Firoz 感兴趣。然而，如果玛丽突然成为名人，人们很快就会对她的心理健康产生兴趣。如果政府突然决定追捕努尔，那么 Skype 可能会在伊朗为她提供掩护，但在沙特阿拉伯则不然 因为 Skype 属于微软，除流氓国家外，微软通常会遵守政府授权令。即使在伊朗，也需要一些 opsec。如果安德鲁使用 Skype 与 Nur 交谈，那么他最好不要使用相同的用户名（或 IP 地址）与他的所有其他皈依者交谈，否则宗教警察会从他们的假皈依者那里得知这一点并上门敲门。

第三个因素是能力，包括支持和动力。在我们所有的用户中，投资银行家伊丽莎白可能是最简单的案例。她的工作是合法的，并且有一个 IT 团队提供支持。如果小心使用，Tor 提供了相当好的匿名性，而且风险很低；如果目标怀疑她的兴趣，

3这个短语也用于黑客，他们根据需要直接利用目标的漏洞来攻击系统，并且不留下远程访问木马。在这种情况下似乎也很合适。

Machine Translated by Google

她只是损失了一些钱,而不是她的生命。格拉齐亚诺面临更高的风险,但背后有一支经验丰富的警察组织。贾斯汀也在玩高额赌注,但管理问题要复杂得多。竞选活动是一场漫长的筹款活动,需要数十名难以约束的志愿者,他们的重点是胜利而不是安全。 Liz 面临重大风险,Irene 提供的支持质量可能会有所不同。 Dai、Firoz、Hristo 和 Nur 在没有任何有能力的技术支持的情况下都面临着极端的危险。

最后是取证问题。我将在后面的第 26.5 节中对此进行详细讨论,但警方面临的主要问题是如今在搜查房屋时发现的数据量庞大:笔记本电脑、手机、平板电脑、相机、电视上可能散布着数 TB 的数据,记忆棒和各种其他设备。如果你不想被发现一根针,那就建一个更大的干草堆。所以 Firoz 的公寓周围可能散布着许多电子垃圾,作为将违禁品藏在加密卷中的记忆棒的封面。并且有许多特殊的方法可以使临时搜索者无法访问内容;他可能会以某种可修复的方式损坏记忆棒,或者只是将其物理隐藏起来。努尔或任何对警方突袭可能是坏消息的人都可能采取同样的做法。

这一切都回到了 tradecraft。什么地方和时间会有所不同,因为这取决于当地对手的实际行动。

为了击败常规的 trac 分析,找一份接待员的日常工作可能就足够了:如果镇上的每个人都打电话给医生的手术室,那么有人打电话给手术室这一事实传达的信息很少。

25.4.4 把它们放在一起

现在回到我们的用户列表,我们如何总结我们学到的东西?

1. 传教士安德鲁承担着最艰巨的安全通信任务之一。他无法会见他的皈依者以在 opsec 中训练他们,并且需要使用一些可用且不显眼的东西。也许对他来说最简单的解决方案是使用 Skype 或 WhatsApp。

2. 就您的女儿贝拉而言,目标是帮助她成长为一个有能力的成年人。我从没想过让我的孙子们使用 Tor;那真是令人毛骨悚然。我所做的是不时在餐桌旁谈论诈骗、网络钓鱼和其他滥用行为。孩子们喜欢这个并慢慢吸收对抗性思维的艺术。这与我们玩的棋盘游戏有着相同的精神。

3. 精神分析师查尔斯应该对风险和可能的缓解措施有基本的认识。当他开始了解他的病人玛丽时,他可能偶尔会提出他认为相关和必要的建议,只要这些建议顺其自然并赋予她力量而不是吓唬她。

   但是,如果这会破坏她对治疗环境的信任,这违背了他所承诺的临床方法,他也可能不愿意提出建议。谈判这种环境可能太难了;由于患者和治疗师之间的不对称权力关系,知情同意是治疗中的一个难题。

Machine Translated by Google

在实践中双方可能都缺乏相关知识,即使 Mary 对风险的了解比 Charles 多,她也可能觉得无法提供任何建议。

4. 人权活动家戴是所有工作中最艰巨的工作之一,但由于她不时受到秘密警察的打击,并与其他可以分享经验的活动家一起工作,随着时间的推移,她可以发展出良好的手艺。

5. 并购分析师伊丽莎白很可能会发现 Tor 可以很好地满足她的需求。她的主要问题是正确使用它并注意她对目标网站的查询类型,以免泄露游戏。

6. Firoz 的情况很糟糕,坦率地说,如果我处于他的处境,我会步行前往德国。如果这不可能,那么他不应该只使用 Tor,而应该买一个 Mac 或 Linux 盒子,这样他就可以减少接触色情网站恶意软件的风险。
他需要提前想好如果他被警察突击搜查会发生什么。 （也许他应该加入革命卫队,这样警察就不会一开始就搜查他。）

7. 格拉齐亚诺也有一份艰苦的工作。在客户端保护一个秘密网络免受一两个叛徒的侵害已经够糟糕的了（安德鲁必须这样做）;在服务器端防御偶尔的背叛更加困难。正如我们在第 10.2 节中描述的那样,他的部分解决方案可能是一个分区的警察记录保存系统,以阻止弯曲的警察访问所有内容。

他还可能使用告密者使用的任何机制与告密者聊天。

8. Hristo 可能会看到使用加密电话的优势,但当警察破解它时,他们可能会卷起他的整个网络。站在他的立场上,我会从 Dai 那里了解到,从长远来看,拥有最佳 opsec 的团队将胜出。所以我会专注于此,并就 trac 安全性对我的同事进行教育。如果我们使用 Signal 等带有临时消息的聊天应用程序,并定期更换手机和 SIM 卡,那么我可以看到我的哪些同事受到纪律处分,并决定信任谁和什么。

9. 记者 Irene 的工作是最具挑战性的工作之一。一名记者不仅需要擅长写故事,还需要善于阅读他人、评估真相和判断风险。调查记者也需要手艺。正如当今的任何记者都需要知道如何使用搜索引擎一样,侦探也需要知道如何保护她的消息来源。对隐私技术有一些基本的了解是不够的;她需要知道如何向可能处于极度压力和生命危险中的联系人传授正确的策略。这意味着不仅要了解人,还要了解威胁和工具。 （随着这项工作变得越来越重要和高技能,媒体可用的预算正在崩溃,因为谷歌和 Facebook 吃掉了他们所有的广告。）

10. 贾斯汀也有一个棘手的问题。很难保护由难以约束且可能有无法改正的不良技术习惯的热心志愿者所开展的短期高后果工作。然而,他可能不了解自己的弱点,只会继续前进,希望获得最好的结果。

Machine Translated by Google

理查德克莱顿写了一篇关于网络空间匿名性和可追溯性的论文,分析了网络匿名性变得多么复杂 [442]。有很多方法,即使是那些没有特别努力隐藏自己的人最终也无法追踪。当骚扰电话来自多人居住的学生宿舍的电话线或预付费手机时,很难确定责任。 ISP 也经常保留不充分的日志,并且事后无法跟踪 trac。但也有很多方式让试图匿名的人失败;最终人们会犯错误,无论他们在 opsec 上付出了多少努力。技术一直在让 opsec 变得更加困难。

这甚至适用于政府安全和情报机构。

25.4.5 姓名邦德。占士邦

2010 年 1 月,我们收到警告,伊恩·弗莱明 (Ian Fleming) 和约翰·勒卡雷 (John le Carré) 的小说中描述的传统情报机构贸易手段开始出现问题。以色列人派出一支由 26 名摩萨德特工组成的团队前往迪拜,以杀死在那里从伊朗购买武器的哈马斯高级官员 Mahmoud al Mabhouh。过去,此类杀戮都是秘密进行的,但这次阿联酋当局收集并检查了所有闭路电视录像,将其与特工的酒店住宿和过境点联系起来。结果发现,其中 12 人使用的是英国护照 其中许多是发给移民到以色列的英国人的,但护照上有特工的照片 还有 6 名爱尔兰人、4 名法国人、3 名澳大利亚人和 1 名德国人。

英国和澳大利亚以护照违法为由驱逐了以色列外交官 [307]。在监控无处不在的现代世界中,边境管制处的生物识别技术和在线护照数据库使得使用假名旅行变得更加困难。

第二次警告出现在 2013 年,当时一份报告分析了 2003 年在意大利绑架一名名叫 Abu Omar 的穆斯林神职人员的事件,并将其归咎于中央情报局,导致一些特工被意大利警方缺席起诉 [1274] .

第三次警告出现在 2014 年,当时中国人从人事管理办公室窃取了整个美国安全许可数据库,如我在 2.2.2 节中所述;这不仅包括整个美国情报界,还包括 2200 万现任和前任联邦雇员。个人信息的武器化仍在继续;2016 年调查权力法案使英国政府能够要求拥有这些数据的公司提供大量个人数据集,从而使这些机构能够访问信用记录、医疗记录等。到 20 世纪末,军方担心中国人正在收集每个士兵的个人信息以用于未来的信息战,而情报机构开始怀疑传统间谍时代是否已经结束 [1274]。国防和情报界以各种方式做出回应,五角大楼告诉员工不要使用消费者 DNA 测试套件,而中国人显然更喜欢低技术含量的东西,比如 dead drops,但尚不清楚是否有灵丹妙药。当几乎每个人都知道这么多时,很难进行秘密行动。

在此背景下,中国争取“人工智能霸权”的举动令人担忧。该国的政治结构鼓励而不是限制这项技术最糟糕的用途:习主席想要一个无所不在的社会控制数字系统,由实时识别潜在异见者的预知算法控制 [48]。

我在 17.3 节讨论了人脸识别;因为中国的城市遍布

Machine Translated by Google

闭路电视系统,他们肯定可以跟踪人们。但这总体上效果如何?机器学习在多传感器数据融合应用程序中的使用并不简单,而且在社会预测方面往往效果不佳或根本无法发挥作用 正如我们在本章前面所讨论的那样。在第 26.4.1 节中,我们讨论了中国制度如何将新疆持不同政见的维吾尔族人口用作测试案例,严重侵犯人权导致美国和欧盟对相关中国公司实施制裁。

同时,正如我们对贾斯汀案的讨论所揭示的那样,在我们更为混乱的民主国家中,很难确保政治运动免受攻击。
Maciej Ceglowski [397] 讨论了 2018 年美国大选导致的运营问题,他还警告了确保选举安全的更广泛问题。接下来我们转向他们。

25.5 选举

正如我在 2020 年所写的那样,继 2016 年俄罗斯干预英国脱欧公投和当年晚些时候的美国大选引发争议后,人们对即将举行的美国大选的行为和可信度感到担忧。

由于投票系统的巨大差异,美国选举多年来一直是投票技术的试验台。已经有很多人试图打败人民的意志,首先是候选人,最近是外部行为者。

我们也有来自英联邦的重要经验,其中包含大多数其他前英国殖民地;其所有成员国都举行某种形式的选举[329]。

选举技术及其安全性的故事是几个世纪以来攻击和防御的共同进化之一。在学校里,我们都学习了现代宪法演变历史的一些变体。长期以来,参与式政府在村庄等小团体层面普遍存在,在村子里,每个人都认识其他人,决策可以通过协商一致或多数人做出;问题是将其扩展到更大的单位,例如城市和州。

希腊人和罗马人试验了选择代表参加议会、议会和法院的机制,但发现民主常常退化为寡头政治,或者君主夺权。他们设计了宪法机制来减少此类失败的风险,包括权力分立、按地理选区而不是部落投票、通过抽签而不是投票选择领主以及任期限制。尽管罗马帝国结束了这些实验,但理想通过瑞士和意大利城邦的教皇选举和中世纪行会得以延续。在英国内战中,议会夺取了国王的权力并砍下了他的脑袋 ; 1689年的定居点使英国成为君主立宪制国家。十七世纪也见证了新世界的第一次集会,导致了十八世纪的美国革命,开国元勋们受到希腊和罗马模式的启发。

其背后还有另一个故事,即享有权力的精英如何不断操纵该系统以保住它。提前选举没有隐私;罗马选民在他们的候选人身后排成一排,公开抗议的投票方式直到 19 世纪仍然是常态,这导致了贿赂和恐吓。这

Machine Translated by Google

英格兰的紧张局势与社会阶层有关:男爵在 1215 年获得了一些权利,随后其他财产所有者在一系列改革中获得了权利。 1832 年的第一次现代改革引入了重新划分选区:在工业革命中兴起的英国城市中,很少有国会议员,而其他选区几乎没有选民,国会议员由当地地主选出。需要一系列改革法案才能将选举权扩大并平均分配给财富和收入依次下降的人,但竞选的高昂成本限制了富人的政治生涯。最终,无记名投票于 1872 年引入。与此同时,在美国,故事更多地与种族有关。内战结束了奴隶制并将选举权扩大到所有人;但在重建失败后,前邦联各州制定了识字测试和其他法律来阻止黑人公民投票。只有在第一次世界大战之后,这两个国家的妇女才被允许投票。

滥用职权的情况很普遍:时至今日,英国、美国和其他地方的政客们都在试图通过公平和不正当的手段让他们的支持者比他们的反对者投票更多。

25.5.1 投票机的历史

从 1800 年代末开始,技术创新浪潮试图阻止美国的选举滥用行为,道格拉斯·琼斯 (Douglas Jones) 和芭芭拉·西蒙斯 (Barbara Simons) 讲述了一个故事 [991]。许多城市和州都有政治“机器”,它们不仅能获得选票,还能操纵选票,利用美国的选举是在州和县一级组织的,而不是像英国那样在全国范围内组织的这一事实。在纽约,Tammany Hall 的 Boss Tweed 有时会填写投票箱,有时只是让他所在选区的工作人员补足结果。为了反击这一点,发明家们想出了从透明投票箱到拉动杠杆时为机械计数器计时的投票机等各种方案。

狡猾的政客和官员适应了。在路易斯安那州,Long 兄弟打败了海豹突击队,使伯爵取得了理想的结果,并统治了该州多年。最终人们意识到,县大楼里维护和编程机器的技术人员控制了结果。机械投票机具有大约 100 位的可编程性,通常以开口销和其他机械联动装置的形式出现,这是其他人无法理解的。磨损也可能包括篡改;技术人员可以通过在相关齿轮上敲几颗齿来导致他们不喜欢的候选人少计。

25.5.2 悬挂乍得

受钢琴演奏者的启发,发明家设计了一种在纸卷上打孔的机器。一旦穿孔卡片因制表机和计算机而普及,它们便得到广泛使用。这个想法是,在卡片上打孔的选票既是人类可读的,又能被机器快速计算。一旦投入投票箱,它也是匿名的(除非你担心指纹)。

在 2000 年的美国总统大选中,结果转向了使用打孔卡片机的佛罗里达州,重新计票涉及到对 chads 的争论 chad 是选民从卡片上打出的长方形硬纸板。是一个“挂

Machine Translated by Google

乍得 ,仍然附在卡上,有效投票?凹坑呢,冲头没有穿透的地方?计票机拒绝了超过 100,000 张选票,而乔治布什对阿尔戈尔的多数票仅为 537 张。最终最高法院停止了重新计票,将选举让给了布什。这引起了争议,以至于国会在 2002 年通过了帮助美国投票法案 (HAVA),该法案拨款 38 亿美元用于购买更新的选举设备。

随着公司争先恐后地向这个巨大的新市场制造和销售机器,淘金热随之而来。这让安全工程师感到震惊。事实上,随着佛罗里达州重新计票的进行,我正在新奥尔良参加应用程序安全会议,与会者包括许多国家安全局和国防承包商的工作人员 ,我们组织了一场辩论。尽管政客们认为应该尽快用电子设备取代机械或纸质投票系统,但安全专家并不同意。绝大多数人以老式举手表决的方式投票表示我们不信任电子选举。国家标准局的 Roy Saltman 1988 年的一份报告已经阐明了大部分可能出错的地方 [1641]。

一些新产品是直接记录电子 (DRE) 机器,是 19 世纪杠杆机器的后代,通常在屏幕上显示候选人和其他选票选项,然后记录选民的输入。后来的研究表明,使用 DRE 机器进行的投票中约有四分之一包含至少一个错误 定义为与选民意图不同的投票。 2006 年在佛罗里达州萨拉索塔广泛报道了这种“倒票”,但尚不清楚根本原因是可用性还是技术(例如,取决于您如何对不敏感的触摸屏进行分类)。无论哪种方式,三分之一的选民都忽略了评论屏幕上的错误投票 [991]。

2002 年选举中报告了许多问题 [806];次年夏天,领先的投票机供应商 Diebold 因安全漏洞将其源代码留在了一个开放网站上。 Yoshi Kohno 及其同事对其进行了分析,发现该设备“甚至远低于其他情况下预期的最低安全标准”:选民可以无限制地投票,内部人员可以识别选民,外部人员也可以入侵系统 [1075]。几乎恰逢其时,活跃于布什总统连任竞选活动的迪堡首席执行官沃尔登·奥戴尔 (Walden O Dell) 写道:“我致力于帮助俄亥俄州在明年将其选举人票交给总统”[1987 年]。这引起了轩然大波,并呼吁通过一项法律来实施 Yoshi 的主要建议,即应该有一个选民可验证的审计线索。 (投票研究员 Rebecca Mercuri 早在 1992 年就认为 DRE 设备应该在窗后的纸卷上显示选民的选择,并让他们在投票前对其进行验证 [1295]。)在一些 DRE 机器中,这在一种记录所有选民行为的非易失性存储盒的形式,但这会造成隐私紧张。其他 DRE 机器根本没有审计线索;审核员所能做的就是要求他们再次打印出相同的结果。

25.5.3 光学扫描

大多数非 DRE 设备由光学扫描机组成,这些机器会扫描选民填写的选票或卡片,无论是用笔还是特殊的选票标记设备,然后将其投入投票箱。光学的

Machine Translated by Google

扫描系统自 20 世纪 80 年代就已问世,是从用于在学校对多项选择题进行评分的标记感扫描仪发展而来的。

在接下来的选举周期中,加州国务卿黛布拉鲍恩授权由加州大学教授大卫瓦格纳和马特毕晓普领导的大型计算机科学家团队对该州的投票系统进行全面评估。这些报告令人沮丧 [306]。他们检查的所有 DRE 投票系统都存在严重的设计缺陷,这些缺陷直接导致攻击者可以利用这些漏洞影响选举结果。

Diebold、Hart 和 Sequoia 之前批准的所有投票机的认证都被撤销,ES&S 迟交的系统也被取消认证。加利福尼亚州可以采取如此激进的行动,因为在 2004 年投票的 900 万人中,可能有四分之三使用纸质或光学扫描选票。

佛罗里达州立大学的科学家对佛罗里达设备进行了类似的检查,他们在 2007 年 7 月报告了 Diebold 设备中的一系列新漏洞 [749]。俄亥俄州效仿并得出了类似的结论。所有被评估的设备都存在严重的安全漏洞:应该加密的数据没有加密;加密做得不好(例如,密钥以明文形式存储在密文旁边);缓冲区溢出;无用的物理安全; SQL注入;可能被篡改的审计日志;和未记录的后门 [1261]。

但是,如果您放弃 DRE 机器以光学扫描纸质选票,就像大多数美国县自 2006 年以来所做的那样,如果接近结果受到质疑,您可以进行人工重新计票。但是仍然有很多事情要出错。

首先,数百个县使用选票标记设备,让选民在触摸屏上做出选择,然后机器打印出一张他们可以目视检查并投入投票箱的投票表。但是有些机器会制作单独的人类可读标记和机器可读标记,如果这样的机器可以被黑客攻击,它可以打印一张选票卡,上面的文字是“戈尔”,但条形码是“布什”。因此,关于您检查的内容以及检查方式有很多细节;最佳做法是设计风险限制审计。在英国,黄金标准仍然是手写选票,但在美国,选票标记机的供应商已经招募了残疾人权利活动家来销售他们的设备。

我们在英国的经历大体上具有可比性,尽管我们从未采用过投票机。托尼·布莱尔政府逐步扩大了邮寄和其他缺席选票的使用范围,这遭到了反对党的批评,因为这让贿选和恐吓变得更加容易。党内工作人员(其中布莱尔的工党人数较多)可以迫使选民选择邮寄投票,然后收集他们的选票表格,填写并提交。将投票从帖子扩展到电子邮件和文本的计划受到批评,因为它使这种现有的低级滥用变得更容易,并可能对自动化开放。最后,在 2007 年 5 月的地方政府选举中,电子投票试点在英国的 11 个地区进行。我的两个博士后在贝德福德选举中担任监票员,观察到与美国各次选举中报道的相同的混乱局面。计数比用纸慢;该系统(光学扫描软件)的错误率很高,导致许多

Machine Translated by Google

比预期更多的选票被发送给人类裁判员进行决定。 （打印机在打印运行的中途更换了墨水，所以一半的选票是“错误的黑色阴影”。）更糟糕的是，该软件有时会将同一张选票发送给多个审裁员，并且不清楚哪个他们的决定被计算在内。最后，为了让大家回家，回国的官员接受了一份保证书（供应商当场写的），保证不会因此而误计选票。然而演习却让各方代表深感疑虑。

组织志愿者的开放权利组织报告说，它无法对观察区域的结果表示信心 [1472]。选举委员会没有反对，这一经验说服了英国，直到今天仍继续使用手工计数、手工标记的纸质选票。 （英国选举滥用行为发生在杀戮链的其他地方，从选民登记到邮寄投票滥用再到违反竞选财务限制：因此修复计算机不足以解决问题。）

25.5.4 软件独立性

这段经历使人们认识到软件独立性的重要性和实现的困难 即投票软件中未检测到的更改或错误不会导致选举结果发生无法检测到的更改或错误的属性 [1608]。我们必须假设计票软件存在错误并且可能是恶意的，因此我们不必依赖它，手动重新计票的可能性是一个重要的缓解措施。但是你如何在实践中做到这一点？

在贝德福德，候选人认为人工重新计票会导致相同的结果，但多数票数不同，并且不想再花 20 个小时进行全面的人工重新计票。

2020 年的共识是，系统的设计必须支持风险限制审计，该审计可以严格限制因软件出现问题而导致欺诈或错误的可能性。对于光学扫描，这可能意味着将来自每个投票箱的所有选票保存在一个单独的包中，以便候选人可以挑战“让我们对第 17、37 和 169 号箱进行手数”，这可以很快完成。如果计数接近或发现差异，您可以手动计数更多箱子。 （事实上，在 2000 年的布什诉戈尔诉讼中出现了关于部分重新计票与全州重新计票的争论。）

密码学家试图使计票更加可验证。对加密选举机制的研究可以追溯到 80 年代初期，当时 David Chaum 提议向选民提供一个数字选票令牌，该令牌使用与数字现金相同的通用技术构建，他们可以将其用于选择的候选人。在第 5.7.7 节中，我描述了该机制：这是一个有趣的密码设计问题，因为您需要同时支持匿名性和可审计性。选民需要确信他们的选票已被正确统计，但为了防止买票，他们不能向任何其他人证明这一点 选票必须是无收据的。

经过三十多年的研究，现在已经有了很好理解的机制。例如，微软研究院的 Josh Benaloh 及其同事的免费 Election Guard 系统允许进行数字选票

Machine Translated by Google

在扫描仪或选票器等选票收集设备中,加密选票可以被计算在内 使用 El-Gamal 加密的同态属性,将两个加密选票相乘与将两个加密选票相乘具有相同的效果明文的。需要做更多的工作来确保所有选票的格式正确并且结果被正确解密,但结果是与软件无关的计数 [223]。这是 2020 年威斯康星州富尔顿在威斯康星州最高法院候选人初选中进行的试点。

加密计票被宣传为“端到端可验证”,但这种说法有些雄心勃勃。它只解决了问题的计票部分。与 18.6.1 节中讨论的电子签名设备一样,您没有值得信赖的用户界面,因此您仍然需要担心选票标记设备或扫描仪中的错误和特洛伊木马。你仍然需要审计。

您仍然需要担心对选民登记、投票簿、结果汇总和结果公布的攻击。而且,如果选票收集设备是选民手机上的应用程序,您就不得不担心像邮寄选票那样的贿选和恐吓。然后你还必须担心手机恶意软件,以及设计和实施的质量。对在美国一些选举中使用过的此类应用程序的详细评估发现了数十个问题 [615]。

25.5.5 为什么电子选举很难

另一个有趣的威胁出现在荷兰。 DRE 投票机在 1990 年代逐步引入,网络权利活动家对此感到担忧。他们进行了一些测试,发现领先供应商 Nedap 的机器容易受到 Tempest 攻击:使用简单的设备,坐在投票站外的观察员可以看到选民选择了哪个政党 [785]。从安全工程师的角度来看,这很有用,因为它导致德国情报人员解密了大量冷战暴风雨材料,正如我在第 19.3.2 节中讨论的那样(Nedap 机器也在德国使用)。激进分子得到了他们想要的政治结果:阿姆斯特丹地方法院取消了所有 Nedap 机器的认证。

至于其他国家,情况好坏参半。在欠发达国家的一些选举中,国家系统地审查反对党的网站并进行拒绝服务攻击;在其他国家(通常是最落后的地区),选举被更传统的方法所操纵,例如提出虚假的刑事指控以使反对派候选人参加投票,或者只是绑架和谋杀他们。世界范围内最好的虐待调查可能是 Commonwealth 的 2020 年报告 [329]。在我写这篇文章时,白俄罗斯大选引发了动荡,“欧洲最后一位独裁者”亚历山大·卢卡申科宣布他在这次选举中赢得了超过 80% 的选票,出口民调显示他的对手斯维特兰娜·蒂哈诺夫斯卡娅(Svetlana Tikhanovskaya)实际上赢得了70%的选票。

他的暴徒强迫她发表让步演说,并将她驱逐到立陶宛,将她的丈夫扣为人质。卢卡申科随后用武力镇压了由此产生的示威[611]。另一个新闻报道是马里总统在一场政变中被推翻,此前有人指控他在五个月前窃取了选举[1200]。

近年也有不少关于人口登记的争吵

Machine Translated by Google

化;在第 7.4.2.2 节中,我描述了欠发达国家如何通过重新发行国民身份证来操纵选举,并使不太可能支持总统的族裔更难获得身份证。即使在登记机制相当健全的地方,如第 17.4 节中提到的拥有 Aadhaar 生物识别系统的印度,当局也可以直接攻击投票权:纳伦德拉·莫迪 (Narendra Modi) 政府于 2019 年通过了一项法律,剥夺了许多穆斯林的选举权,尤其是边境地区的穆斯林.

这是一本非常古老的剧本。正如我已经提到的,直到二十世纪,英国的选举历史都是关于穷人是否可以投票,而在美国则是关于黑人是否可以投票。即使在 2000 年的佛罗里达州,由于滥用登记而被剥夺选举权的选民人数也超过了因挂票而引起争议的选票人数。正如政府可以通过让没有汽车的人更难投票来影响选举一样,如果你没有电脑,它也可以使投票更难。关于选票或投票机是否过于复杂以至于剥夺了受教育程度较低的人的选举权,也引发了诉讼。

一些关于技术安全的争议已经告上法庭。例如,在我写的 2020 年,佐治亚州看起来一团糟;在多年试图增加投票难度、未能修复 Diebold 机器的已知缺陷并成为俄罗斯人的目标之后,州政府被法院命令更换其系统。新系统在 2020 年 6 月的初选中崩溃,无法满足选民的需求 [851]。

然而,注意力的主要焦点已经转移到社交媒体在选举中的使用。巴拉克奥巴马在 2008 年和 2012 年有效地使用了 Facebook,促使其他人学习社交媒体; 2016 年的选举落到了唐纳德特朗普手中,他不仅在使用 Twitter 方面比希拉里克林顿熟练得多,而且最终为他的 Facebook 广告支付的费用也少得多。正如我在第 8.5 节中解释的那样,Google 和 Facebook 使用的广告拍卖机制将您的出价乘以一个称为“广告质量”的因素,这是人们点击广告的可能性,对于社交媒体而言, 分享它。结果是极端主义偏见:煽动性广告更便宜。

正如我在第 2.2.3 节中所述,2016 年的另一个因素是俄罗斯的干涉。俄罗斯特工不仅为特朗普竞选,经营巨魔农场和旨在压制黑人选票等社交媒体广告活动;他们入侵了克林顿竞选主席约翰波德斯塔的 Gmail。

他们侵入了伊利诺伊州和佛罗里达州（可能还有其他一些州)的系统,并可能操纵选民登记,但他们选择不触发这些攻击,因为他们不需要它们;他们反而攻击了选民。如果克林顿获胜,那么如果这些州中的任何一个投票给她，“欺诈”的证据就会出现,从而破坏她的总统任期。

这一切将如何影响将于 2020 年 11 月举行的选举?由于这本书即将出版,我只想指出,在爱荷华州的民主党初选中,结果聚合已经出现了惨败 [637],而俄罗斯人再次在网上开展煽动性的亲共和党活动 [1619]。

Twitter 和 Facebook 都删除了特朗普及其同事发布的包含有关 Covid [1030] 虚假信息的帖子,人们担心他或其他人可能会利用在线媒体破坏选举进程或信心

Machine Translated by Google

在结果中。特朗普在共和党全国代表大会上做了铺垫,声称只有在选举被窃取的情况下他才会输。 Facebook 内部存在一种焦虑,即尽管扎克伯格表示他将阻止压制选民的企图,但他一直在让右翼更轻松 [1740]。 8 月,主要科技公司宣布结盟打击选举操纵 [963]。

但是事后对结果有争议怎么办?一个多世纪的美国政治史警告我们不要寻求政治问题的技术解决方案。

在欧洲不同的政治文化中,我们有着长期的竞选资金限制传统(美国在最高法院的公民联合会决定将其变成一场混战之前也是如此)。政党在每个竞选活动和每个候选人上只能花费这么多;大多数欧洲国家禁止在竞选期间投放付费电视广告。但执法力度一直在稳步减弱。例如,在英国脱欧公投期间,两次脱欧运动都超过了支出限额,但刚刚支付了 20,000 英镑的最高罚款。俄罗斯对英国退欧的参与主要以财政捐助和社交媒体上的进一步竞选活动的形式出现。可以采取什么措施来阻止此类滥用行为?

在 Open Rights Group 的 2019 年会议上,我主张我们应该将广告禁令从电视广告扩展到 Facebook、Twitter 和 YouTube 上的所有广告。这不仅是为了避免美国最糟糕的大手笔政治,还因为针对个人而非所有人的政治广告助长了极端主义和支离破碎的政治话语。政治家的工作是调解社会不同利益相关者之间的冲突;如果这些团体最终陷入他们自己的过滤泡沫中,那么我们的政客可能会被引诱去激化冲突。禁止广告不会是万灵药(印度在 2019 年禁止了 Facebook 广告),但它将使选举竞赛更多地保持在欧洲人熟悉的文化和经济空间内。

选举仍然是棘手的安全工程问题之一。虽然个别问题 例如选民登记、投票、计票、结果汇总和审计 都有相当稳健的解决方案,但将它们组合成一个稳健的系统并非易事。用于登记选民、记录选票和计票的计算机系统具有许多特性,这使得它们几乎成为稳健设计、实施、测试和部署的病态案例。首先,选举日期是固定的,无论是否准备好,软件都必须在那时部署。其次,不同的地区和国家有不同的要求,并且随着时间的推移而变化。第三,在两次选举之间的漫长间隔中,有经验的员工离职,专业知识流失。第四,必须更新操作系统和其他软件以修复已知的漏洞,更新也可能以无法预料的方式破坏安全; Windows 更新导致 EV2000 投票机将上一个选民的选择突出显示给下一个选民 [991]。然而,美国使用的大多数投票机已不再生产,那么更新从何而来,又将如何进行测试?最后,选举是高压力事件,这增加了出错的可能性 [1357]。

现在让我们从工程上看政治。如果发生攻击,获胜者不想调查可能出了什么问题,如果他们可以避免的话正如我们在 20164年在美国和英国所看到的那样。4正如我所写,诉讼仍在继续,试图强制发布的编辑部分

Machine Translated by Google

选举的“客户”是失败的一方,在没有任何补救希望的情况下 无论是通过法庭,还是通过下一次的投票箱 对民主机制的信任可能会开始失败。但是没有“设计者”来确保机制和法律在整个选举周期中始终保持一致。

相反,通常是在任者调整法律,购买投票机,并在当地政治文化所能容忍的范围内为自己的一方(无论大小)创造尽可能多的优势。虽然投票机制可以支持民主共识,但它们无法取代民主共识:有太多其他方法可以破坏结果。如果潜在的社会契约受到侵蚀,超党派环境会使现任者感到他们不敢放弃权力。在最坏的情况下,结果可能是内战和失败的国家。

25.6 总结

2020 年一些最具挑战性的安全工程问题与这样一个事实有关,即随着软件在我们使用的服务和我们周围的设备中变得无处不在,这些服务和设备的设计面临着人类社会潜在的复杂性。我们看了四个例子。

自动驾驶汽车可以应对空荡荡的沙漠道路,但要找到真正的人类司机则要困难得多。机器学习机制只能走这么远;他们可能在模式匹配方面很出色,但缺乏理解,这为堆栈中各个级别的滥用开辟了新的可能性 尤其是当人们急于将它们用于他们本质上不适合的社会预测任务时。增强隐私的工具和技术是探索人类复杂性的安全后果的一种方式,但无论我们如何努力加密和匿名化事物,社会结构往往会通过这种或另一种方式显示出来。最后,我们进行选举;当现任统治者准备做他们认为可以逃脱的一切 无论是在法律范围内还是在法律范围之外 以继续执政时,我们可以了解到很多关于技术和法律的局限性。

随着越来越多的人类生活转移到网上,在线应用程序的重要性和复杂性也在同时增长。许多熟悉的问题一次又一次地以越来越难以处理的形式出现。传统的软件工程工具帮助开发人员在系统复杂性下降之前进一步攀登 。什么样的工具、技术和治理流程适合处理现实社会的复杂性?这如何与政治互动?这些是我们将在本书的第三部分尝试解决的主题。

研究问题

一大堆研究问题是围绕如何在人与自动化之间分配责任。 HCI 大师 Ben Shneiderman 认为,人为控制加上广泛的自动化是系统可靠、安全和值得信赖的最佳点 [1723]。这对于飞行控制系统和生命支持系统来说是很自然的

穆勒报告那次选举。

Machine Translated by Google

机器,但将其扩展到推荐系统和仇恨言论检测之类的东西并非易事。人类如何对大型科技公司每秒做出的数百万过滤决策进行质量控制?在此基础上的治理应该是什么样的?这一切的背后是关于自动化（包括 ML）是否正在走向人工智能或智能增强的长期争论 [87]。

随着涉及 ML 的自动化变得越来越普遍,问题可能会变得更广泛。建筑师和城市规划者将不得不考虑我们如何设计必须考虑到多个利益相关者利益的生活和工作环境。然后,围绕机制和社会的共同进化将出现全球性的社会和政治问题。在第二版中,我说过“从现在到本书第三版之间的一个关键研究问题……将是保护机制如何扩展……如何发展 安全 （或任何其他紧急属性)在拥有数十亿用户的社会技术系统中。”我注意到简单的规则系统,比如政府钟爱的多级安全,从来都不是自然的,人们总是不得不打破它们来完成他们的工作。那么还有什么?

我们现在有更多的经验;几家大型科技公司运行的系统拥有超过 10 亿的活跃用户,数百家公司的活跃用户超过 1 亿。

在这样的系统中,技术和行为相互适应,但系统开发人员更强大并且有不同的动机（他们想要数据,而用户想要隐私）。人类进行大规模规则谈判的基本机制是市场竞争和政府监管。仅凭这些都不够,技术与政治之间的相互作用甚至可能破坏政府选举机制。

我们工程师需要关心这些问题,并努力理解它们。
在本书的第三部分,我们将尝试解决更广泛的政策和管理问题（例如监视与隐私）,如何管理和治理大型复杂系统的演变,以及如何监管技术以满足社会目标比如安全和隐私。

进一步阅读

对于汽车安全的介绍,您可能首先查看查理米勒和克里斯瓦拉塞克关于他们如何入侵吉普车的描述 [1316],然后如果您想深入了解技术细节,请查看克雷格史密斯的“汽车黑客手册” [1792] .

Nicolas Papernot 的“Marauder s Map”可能是目前快速发展的对抗性机器学习领域的最佳介绍[1493],而 Gary McGraw 及其同事提供了设计原则以及在处理安全性时需要考虑的事项列表具有机器学习组件的系统[1267]。谷歌的机器学习高级副总裁 Je Dean 在 [528] 中描述了该公司对 AI 公平性的研究。我自己关于 AI 与 IA 辩论的哲学立场可以在 [87] 中找到。

至于面对敌对国家行为者的个人隐私,随着双方工具的发展,这是一场不断变化的冲突。一个起点可能是 EFF 网站上的“监视自卫”页面 [618]。有个有趣的

Machine Translated by Google

Ben Collier 对 Tor 项目的组织和社会动态的描述,该项目维护着领先的在线匿名服务,见 [458]。有关更多技术深度,请参阅有关 Tor 的第 20.4 节或 [125] 中的匿名参考书目。

道格拉斯·琼斯 (Douglas Jones) 和芭芭拉·西蒙斯 (Barbara Si mons) 讲述了美国投票制度的历史 [991]。美国国家科学院、工程院和医学院针对 2016 年的事件制作了一份关于选举安全的广泛报告 [1388]。最近,英联邦根据其成员国非常多样化的经验制定了选举电子安全指南,还涵盖了从登记到投票、计票和结果传达的整个周期 [329]。