青未了 新知周刊

A14—15
齐鲁晚报
2022年2月25日 星期五

主编:任志方 责编:于梅君 美编:马秀霞 组版:侯波

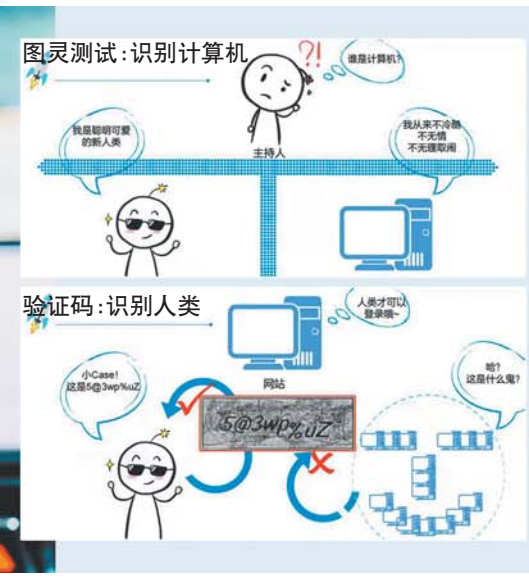
急匆匆地购买火车票,输入账号密码后,跳出一幅九宫格要求点击指定图片验证,好不容易玩完"大家来找茬",一看火车票已被抢光,这样的情形你是否遇到过?对,绊住你的就是网络时代的老朋友——验证码。

据估算,全世界网民一天要输入近2亿次验证码,按照每次输入花费十秒来计算,每天花在验证码上的时间已超过50万小时。

既浪费了时间,体验性又差,那么,验证码存在的意义在哪里?它"验证"的原理又是什么?

齐鲁晚报·齐鲁壹点记者 于梅君

我们每天都要和验证码打交道。



地球人每天被验证码"偷走"50万小时

——扒一扒验证码背后那些不得不说的秘密

1 验证码的诞生:证明你是一个人

"在互联网上,没人知道你是一条狗"——漫画《纽约客》中的这句名言,曾被视为网络自由的宣言。对于网页和应用的所有者来说,很多时候确实无法识别网络的另一端到底是机器还是人。

"现在很多网站注册和登录都要用到验证码,就是为了区分计算机和真正的人。"南京大学信息科学博士、南京视网么信息科技有限公司创始人张帅介绍,验证码英文"CAPTCHA"直译就是"全自动区分计算机和人类的图灵测试"。

所谓图灵测试,是人工智能圈一个著名实验,如果测试者向无法确认身份的两个对象(一人、一机器)提出相同的一系列问题,得到的答案让他无法区分究竟谁是机器、谁是人,那么则认定机器通过测试。也就是说,"图灵测试"是由人类来判断:谁是计算机?谁是人类?

验证码则是图灵测试的反向和变种,就是由计算机来判断:谁是人类?谁是计算机?最重要的,是识别出人类。

早期网站登录都是依据用户名与密码,但黑客程序有可能针对某一个特定用户账号采用穷举破解的方法,不断进行登录尝试,造成潜在威胁,例如:刷评机器人用大量垃圾评论和广告淹没真人用户留下的有价值信息;在票务网站,就算你有三头六臂也抢不过自动刷票软件;网站被人一次性注册几百万个垃圾账号,产生垃圾信息或者操纵投票……因此,各种计算机系统都需要一个坚固的"盾"来保护自己,于是验证码应运而生。

验证码出自美国卡内基梅隆大学研究人员的设计:计算机程序难以识别手写的文本,而人类可以轻易看懂,于是程序员在注册账号时设置了一道门槛——必须输入"歪曲"的文本才能完成注册,防范那些可能对在线服务造成威胁的自动程序,如恶意破解登录密码、刷票、论坛灌水、刷网页等。

北京邮电大学信息与通信工程学院教授牛凯说,理论上,只有真人才能通过推理分析验证码图片中的字符。隐蔽在杂乱背景中的扭曲字母,人眼往往可以准确辨识,而采用计算机,识别准确率较低。

验证码的影响力之大,已扩展到全球范围——每天有大约2亿次的验证码输入,每个验证码平均耗时大约10秒钟,那么全人类每天花在输入验证码上的时间就超过50万个小时。

2 形形色色验证码:『矛』与『盾』攻防战

"有了图形验证码,有效避免了暴力请求破解的威胁。在图形(数字)验证码的基础上,慢慢演化出滑块验证码、语音验证码、点击型验证码、图像验证码、短信验证码、生物特征验证码等形式。"张帅说,短信验证码可用于对安全性要求较高的应用,比如支付宝、登录银行客户端等,通过运营商来发送短信,安全上有保障。每个验证码与手机号码相对应,而且有效期仅为60秒—90秒,可在一定程度上避免账号密码泄露、身份伪造等行为。

点击型验证码的最大特点是,使用者只需点点鼠标就行,通过这种人类专属的行为动作,以及使用者在浏览器中的一些操作数据、浏览数据等,识别出真正的人类。在一些重要的计算机系统,如动车售票网站、大型购物网站、大型视频网站等用户量较大、数据安全要求高的地方,采用点击型验证码,可以有效避免其他计算机攻击。

滑动型验证码通过收集使用者的动作,判断是否为人类。例如:人拖动滑块的轨迹会是一个先快后慢的过程:先快速拖动,后慢慢对齐,再瞬间释放。这种验证码的用户体验较好,虽然被计算机技术突破的成功率有60%以上,但计算机技术模拟人类行为的成本较大,得不偿失。

各种人脸识别、指纹识别、声纹识别甚至虹膜识别,都算是生物特征型验证码,由于其独一无二的特性,犹如一把专属钥匙,具备较高的安全性。

层出不穷的"计算机技术"不断提高自己伪装成人类的能力,而千奇百怪的"验证码"则保护着各种计算机系统。"比如,黄牛不可能人工去买票,必然是使用抢票软件。"北京邮电大学信息与通信工程学院教授牛凯说,黄牛可能手中囤积大量身份证号码,刷票软件可以自动登录,比正常的手工操作快几十倍乃至上百倍,因此可以大量刷票。

中国铁道科学研究院电子计算技术研究所副所长朱建生也表示,不用图形验证码,机器抢票时间为0.1秒/张,人工抢票则为2秒/张,而使用图形验证码,由于机器无法自动识别,令票贩子无法再利用刷票软件囤票倒票。

可见,如果没有验证码,大批用户的数据安全将成为黑客肆无忌惮的攻击对象。而自己收到的各类验证码,是信息、资金的最后一条安全防线,为防止被骗子趁虚而入,最好记住一句话:"验证码打死也不能告诉别人!"

3 AI越来越聪明:验证码将何去何从

在人工智能不断发展的现在,机器能通过越来越多类型的图灵测试。我国西北大学房鼎益、陈晓江教授团队研究发现,网站上看似复杂的文本验证码存在"巨大安全漏洞",大多数可被人工智能破解。

房鼎益教授介绍,团队基于最新的人工智能技术,建立了一套新型验证码求解器。他们综合分析了全球最热门50个网站的文本验证码,包括公众熟知的谷歌、微软、淘宝、百度、腾讯、京东等网站。

实验证明,大部分文本验证码可在0.05秒内被人工智能攻破,大部分网站文本验证码的破解率能达到50%以上。"验证码一旦被攻破,写个程序就能成为水军,用机器点赞或投票;也能刷抢火车票,这是人工操作做不到的。"房鼎益说。

西北大学信息科学与技术学院副教授汤战勇说,通过这项研究,希望能提高业界对文本验证码安全性的重视和关注。

如果AI学会识别验证码并被别有用心的人利用,有哪些反制措施?汤战勇认为,不管AI多聪明,验证码都不会被淘汰。他说,没有绝对安全的系统。验证码和反验证码技术会在此消彼长中前行。破坏安全的方式越来越刁钻,安全措施也会越来越严谨,不必太担心。

在视觉图像领域工作多年的系统架构师王之琳表示,验证码已进入智能时代,很多操作变得更简单,用户只需在页面上点击"I'm not a robot"(我不是机器人)的勾选按钮即可。

其实从用户打开页面,加载出验证码的那一刻起,校验过程就开始了。通过用户在页面上的停留时间、鼠标移动速度、位置偏移等作为参考,将这些复杂数据传到校验服务器的后台进行AI分析,就能判断是不是真人操作。

房鼎益教授表示,目前研究人员正致力于利用人工智能合成更安全的验证码来抵御攻击。"我们试图在不影响交互性的基础上,让用户体验更便捷,让机器更难识别,确保网络安全和用户隐私不被泄露。"

延伸报道 全世界都在为验证码免费打工?

花样繁多的验证码不断问世,难道只是为了安全所需?不,还有另一个隐藏目的:让用户免费为其打工,比如,应用于纸质书籍电子化。

验证码的发明者易斯·冯·安说,他的创造其实无意中浪费了人类最重要资源,具体来说,这种验证码每天都会让大家看2亿个单词,每个单词大约花费10秒钟,也就是每天会浪费大约50万小时的人力资源。那能不能把这些被浪费的资源利用起来呢?他发现,许多公益组织正在把旧书籍扫描成电子版,但对计算机来说,那些斑驳的文字太难识别了。书籍内容大部分是文本,验证码也是文本,把扫描版的书籍文本对接到验证码上,让用户来识别不就行了?于是,2007年路易斯推出新的验证码系统reCAPTCHA,它会提供两个单词给用户识别,这两个单词都是书籍扫描版的一部分。

计算机其实已识别出第一个单词,之所以要展示出来,就是为了测试一下用户是不是真人,不过第二个单词计算机无法识别,用户需要录入自己认为的结果。面对第二个单词,一旦有10个人输入了同样答案,那么这答案就会被当作正确答案。也就是说,真正有效的人机测试,在验证码的第一部分就完成了,而第二部分,则是用户在义务为人类文明做贡献。

粗略统计,现今全世界每天都有2亿个字符通过reCAPTCHA录入,相当于人类15万小时的工作量。在不知情的情况下,全世界用户每年免费将230万本书数字化,并在短短几周就把CNN和《纽约时报》数年累积的内容翻译成其他语言。

除了义务为人类文明做贡献,输入验证码其实也在为人工智能贡献着一份力。小伙伴们肯定遇到过街景验证码,让我们选中小轿车、路牌或是自行车等图片。在你费眼又费脑地输入它们时,你其实是在为谷歌的人工智能免费打工。因为这些图片大都来自谷歌街景,其中部分图片是AI已经识别出来的,用于识别你是否真人。

和之前的书籍电子化一样的套路,其中还夹杂了几张AI难以识别的街景。从用户这里搜集到的街景数据,帮助训练人工智能,使得人工智能可以像人眼一样准确地识别路况信息。

在用户的"辛勤喂养"之下,运用了谷歌AI技术的无人驾驶汽车Waymo,已经在自动驾驶领域处于遥遥领先的地位,被称作是世界上最可能先达到L5级别(完全自动驾驶)的公司。

如今,收购了reCAPCHA的谷歌,已把大部分验证码都升级了,用户只要点击一下"我不是机器人"的按钮,就能轻松通过验证。即使这样,验证码还是能从你身上吸点油水。因为在你点击按钮的同时,鼠标的运动轨迹甚至是你打开的网页都可能被收集,帮助验证码系统进化。