智能社会研究

（双月刊）

2024 年
第 1 期
2024 年 1 月 10 日出版

2022 年 11 月 10 日创刊

总第 8 期

目　次

论文

研究报告

译文

书评

访谈

CONTENTS

THESIS

RESEARCH REPORTS

TRANSLATED TEXT

BOOK REVIEW

INTERVIEW

数字时代安全科技价值报告*

陆菲菲　秦　昊　张友红　吕　鹏**

摘要：本文基于2023年全球数字社会的安全风险新形势和科技创新内在逻辑，对安全科技做出“公共品”的新定义，总结出其具有的压舱石与助燃剂双重价值，提出人工智能和安全科技将会是未来的两项通用技术。政企合作是安全科技成为公共品的核心驱动力，通过更多的公共力量投入，构建更好的安全科技实力和制度，采用数字化转型的“慢起飞模型”，新的安全技术将得到更好的发展，个人与社会为数字化转型所付出的代价也将更小。

关键词：数字时代　安全科技　AI安全　大模型安全　网络安全　数据安全

一、前言

安全已经成为数字时代最为重要的话题，我们正处于安全风险的质变时刻。随着数据爆炸式增长，数字化基础架构错综复杂，网络攻击面不断扩大，安全风险也呈现出新情况。尤其是在信息安全方面，近年针对数据的攻击、窃取、滥用等各种风险事件不断涌现，对经济、社会产生了巨大影响。

2023年人工智能（AI）大模型横空出世，大大推进了深度智能时代的到来，然而它对安全局面的影响同样可能是颠覆性的。最近引起全球关注的OpenAI公司因对AI安全的不同观点而导致的人员变动风波，很好地说明了业界对当前

* 本文系中国社会科学院2024年度实验室孵化专项资助“社会计算与数字社会前沿研究”（项目批准号：2024SYFH010）的阶段性研究成果。感谢中国社会科学院大学哲学院教授、中国社会科学院科学技术和社会研究中心主任段伟文，科技部“云计算+大数据”国家重点专项课题负责人、数秦科技有限公司执行总裁兼首席科学家崔伟，北京大学客座教授、蚂蚁集团首席技术安全官韦韬，浙江工业大学网络空间安全研究院院长宣琦等多位专家学者在本文写作过程中提出的宝贵观点和建议。

** 陆菲菲，蚂蚁集团安全实验室。秦昊，蚂蚁集团安全实验室。张友红，蚂蚁集团安全实验室。吕鹏，中国社会科学院大学数字中国研究院。

AI 安全高风险的态度，以及它在整个深度智能社会中投下的令人不安的阴影。因此，站在深度数字化的转折时刻，对安全科技的关注与投入已刻不容缓。

安全科技的价值，既体现为压舱石，更重要的是作为助燃剂：压舱石指的是守住技术的安全底线，防御可能存在的风险隐患，支撑和保障数字社会的稳定运行；助燃剂是提高技术的安全上限，降低技术运行的成本，让新技术得以规模化落地，推动产业焕发新机，让更多人从中受益。

数字化为何有安全风险？其实，风险自古至今都存在，只是因为产业运营、个人生活、国家组织的行为模式向数字化转型，风险也就相应地并轨转移。隐私数据泄露、商业勒索、基础设施瘫痪……放眼未来，有新科技的地方，就会有新的数字化安全风险，两者是本体和阴影一样不可分割的伴生关系。

我们在前段时间经过综合调研，同时结合多位专家学者提供的宝贵观点和建议，撰写了这份报告。本报告主要关注两大议题：第一，在安全风险质变时刻，更好地理解数字时代的安全风险新形势和科技创新的内在逻辑，提高社会对安全科技的认知、重视与参与度，这是本报告的核心议题；第二，安全科技行业存在多种需要迫切解决的问题，尤其是缺乏明确的生态标准来加快前沿技术的规模化落地，缺乏合理的市场化机制及生态联动方法推动安全科技健康发展，这也是本报告关注的关键问题。

本文的主要结论是安全科技将成为社会的公共品，是数字社会的优先事项之一。未来通用的技术有两个，一是人工智能，二是安全科技。人工智能作为最具颠覆性和战略性的关键核心技术，将在各行各业数字化的基础上进一步实现智能化，成为未来生产力的发动机；安全科技将作为方向盘，始终将各种新兴科技控制在向善的道路上，它保护的不只是若干行业、某个业务、几段数据，而是这个社会里的每一个人，是技术向善的关键保障。安全科技与人工智能技术并举，双方的融合创新与发展，将是我国强国战略中不可忽视的重要助推因素。因此，这两个领域，应该成为我们最关注、投入最大的领域。

我们希望通过这份安全科技价值的研究报告，能更加深刻地洞悉安全科技的未来发展之路，提高社会对安全科技的认知，实现安全科技与人工智能的并举、融合与创新，助力经济社会高质量发展。

二、数字社会面临的全新风险挑战

（一）数字化安全风险的三个历史阶段

数字化安全风险的历史与信息技术的进步保持同步，并非最新的产物。它长久以来处于公众雷达之外，本质原因是 20 世纪末 21 世纪初我国上网民众基数不高。但在数字化突飞猛进的今天，数字化安全风险日益引发关注。我们可以把数字化安全风险的演变划分为三个阶段。

第一阶段是互联网“上古”时期至 20 世纪 90 年代初，这是网络安全风险的萌芽时期。事实上，前互联网时代就已经有了网络攻击，因为保护位于计算机内部的数据也属于网络安全的范畴。1987 年，第一个防病毒软件产品出现。随着互联网的民用化程度提高，越来越多的人开始上网，网络犯罪便开始有利可图。20 世纪 90 年代，网络安全风险暴发，早期那些炫技的个人黑客逐渐消失，黑客活动转向牟利驱动，形成了如今仍在不断壮大的网络犯罪产业。

第二阶段是 20 世纪 90 年代中后期至 21 世纪的最初 10 年，这是计算机病毒及网络黑客的暴发时期。这一时期，互联网高速发展以及社交网站、网络游戏、电子商务日益繁荣，大量计算机病毒涌现，各种网络攻击手法不断出现并呈暴发式增长，网络诈骗已具雏形。21 世纪初，在美国等互联网较早普及的国家，网络犯罪的规模、形式和恶性程度都达到了新的高度。例如，在一次网络攻击事件中，黑客获得了超过 30 亿用户的雅虎账户访问权限。2002 年 1 月，美国记录了第一起网络犯罪谋杀案，一个罪犯通过雇用黑客入侵医院的计算机系统，更改了他的竞争对手给病人开的处方，错误的处方最终导致一名病人死于一名无辜护士手中。

第三阶段的主要特征是数字化安全风险更快迭代、高智能、全覆盖。近 20 年来，网络黑色产业链日益成熟，信息泄露、网络诈骗等网络威胁迅速大范围流行，网络犯罪的规模、范围、严重程度和复杂性均呈指数级增长，且日益超越地理边界，极大地挑战了传统司法系统应对挑战的效力。2021 年 2 月初，一个黑客远程访问了美国佛罗里达州一家水处理厂的计算机系统，试图将供水中的氢氧

化钠含量增加到潜在的危险水平;2022 年 1 月,当国际油价已创近 7 年新高时,因遭到勒索软件的攻击,荷兰与比利时的几处港口无法完成石油装卸和转运,多艘油轮无法靠港;2022 年 4 月,保加利亚的国家邮政系统遭到网络攻击,柜台养老金发放业务被迫中断。

网络犯罪之所以能造成如此大的破坏,是因为它如今是一个收入颇丰的“产业”,拥有机构化的管理结构和研发预算。攻击者能够使用 AI 等高级技术工具,并能够持续加快攻击生命周期,从侦察漏洞到发起攻击,仅需几天乃至几小时。这些例子清晰地表明,数字化安全风险已经进入快迭代、高智能、全覆盖的新阶段。

在这个全新阶段,对抗安全风险已经远远不是商业机构内部的 IT 工作,而关乎每个人的幸福指数和全社会的安全稳定。2021 年 1 月,美国国务院批准设立新的网络空间安全和新兴技术局(CSET);同年 3 月,美国政府发布《国家安全战略临时指导方针》,将提升网络安全作为首要任务,并鼓励私营部门与各级政府合作,同时加大基础设施投入,保障美国免受恶意网络活动侵害。2021 年,我国通过了《国家安全战略(2021—2025 年)》,强调加快提升生物安全、网络安全、数据安全、人工智能安全等领域的治理能力,将安全提升到国家战略层面,实现国家安全战略与国民经济和社会发展规划同步。

(二)AI 时代数字安全风险的新挑战

2023 年 AI 大模型横空出世后,全球业界迅速形成共识:AI 安全风险陡增,这是全世界需要迅捷反应、共同面对的巨大挑战。目前来看 AI 安全风险主要可以分为三类:

第一是内生安全风险。从 AI 内部视角看,AI 系统和一般信息系统与应用一样,由软硬件系统组成,会存在技术本身的脆弱性、对数据的依赖性等自身缺陷导致的安全问题。比如,AI 过分依赖数据与算法,如果给数据库不断投喂带有特定价值观的数据,就会对 AI 系统产生严重的干扰,很容易产生数据偏见、观点霸权等问题,其中就涉及业界广泛关注的 AI“价值对齐”与“决策黑盒化”问题。“价值对齐”就是应确保 AI 的价值观跟人类的价值观相匹配,确保 AI 以对人类和社会有益的方式行事,不对人类的价值和权利造成干扰和伤害。目前来

看，仅靠 AI 大模型技术本身不太可能实现这个目标，需要有更多其他技术的介入。“决策黑盒化”是指难以彻底理解 AI 系统内部的结构和运转，只能通过其输出的内容来判断并得到一种外在认知。未来，如果我们真的要大规模利用大模型的能力，就必须解决黑盒问题，否则 AI 的安全永远是无根的。

第二是衍生安全风险，即 AI 系统因其自身的脆弱性被利用或被不恰当使用，从而引发其他领域的安全问题。比如，AI 系统失误可能引发一些重大安全事故。随着 ChatGPT 等 AI 大模型技术的成熟，某些别有用心的人将其作为违法犯罪活动的工具，如生成虚假新闻，深度合成伪造进行诈骗与钓鱼攻击，侵犯他人肖像权、隐私权，AI 大模型生成的内容本身也可能带有偏见。AI 衍生安全问题影响 AI 的合规使用，还涉及人身安全、隐私保护、军事与国家安全、伦理道德和法律规范等一系列与社会治理有关的挑战性问题。

第三是外生安全风险，也就是面向 AI 系统的外部入侵攻击。当前，安全风险变得更加复杂、隐蔽、强对抗和更具破坏力。一方面，数字化、网络化程度越高，可被攻击的部位、节点就越多；对数字技术的依赖程度越高，网络破坏造成的后果就越严重。另一方面，黑灰色产业链攻击模式也随着业务创新、新科技的引入，逐渐呈现专业化、多样化、智能化、产业化的特点。因而，如何将 AI 驱动的业务风控系统建设得更强更智能以应对大规模网络攻击与入侵，已成为各个企业健康发展的必要课题。

出于对 AI 大模型风险的担忧，2023 年 3 月，包括“深度学习”教父班吉奥（Y. Bengio）、特斯拉 CEO 马斯克（E. Musk）在内的多位科学家和业界顶级专家，参与签署了一封名为“Pause Giant AI Experiments：An Open Letter”的公开信。信中呼吁，因为 AI 实验室陷入了一场失控的竞赛，致力于开发和部署更强大的数字思维，但是却没人能理解、预测或可靠地控制这些大模型，所以所有 AI 实验室应该立即暂停训练比 GPT-4 更强大的 AI 系统，为期至少 6 个月。

这些问题离我们并不遥远。即使我们搁置 AI 有朝一日是否会脱离人类控制等由于“黑盒”而导致的更“哲学”的争论，在当前阶段，StableDiffusion、Midjourney 等 AIGC 工具已展现出了以假乱真的生图能力，ChatGPT 等文本生成模型也已具备高度接近人类的语言能力，一旦被别有用心的人使用，当前并没有足够有效的办法对其进行辨别与控制。随着这些 AI 技术的推广和进一步成熟，其

破坏能力将远超当下的网络欺诈等行为,给社会造成大量的混乱,给个人带来精神和财产损失。考虑到 AI 进步的速度之快、影响范围之广,AI 安全风险不可谓不紧迫。

此外,仅仅一年多时间,市面上已经涌现了各种各样的大模型。这些大模型之间如何更好地协作,共同演进而不是各行其是造成资源的浪费与社会运行的混乱,或者在协作中触发隐私保护等问题,也是未来的关键挑战。

(三)新技术与新安全风险的伴生

除了上网人数快速增加这一诱因,近 20 年网络犯罪的急剧增长,也与科技进步有关。可以说,数字化安全风险是科技发展的内在组成部分,即使互联网行业消失,所有社交媒体全部下线,安全风险依然会存在,并成为社会的潜在威胁。因为数字化并非互联网及相关行业的特有属性,而是几乎涉及所有行业、所有新兴技术。

“工业 4.0”也称“智能制造”“第四次工业革命”,即通过将信息技术集成到运营技术中,来提高工业生产的表现。这些信息技术就是我们耳熟能详的 5G、云计算、人工智能、XR、数据科学等。从全球来看,有两大因素驱动制造业的数字化转型:一是越来越多的制造型企业希望跳过中间商、服务商、代理商等,与用户直接交互;二是制造型企业希望从销售产品转为销售服务,譬如车企。

新技术工具改变了工业的运作方式,同时也导致了安全风险的蔓延,包括网络被入侵攻击,导致服务瘫痪、关键信息和数据被窃取、大范围的人身伤害。

以供应链安全为例,供应链安全风险分为两种,一种是网络遭到破坏,数据中心停止工作,可能导致整个供应链中断或瘫痪。如果遭到中断或瘫痪的是一些涉及国计民生的重要物资供应链,就会造成很大的社会影响。更危险的是,可能会发生数据篡改,导致供应链出错,譬如将不正确的药品发给了某一人群,最终造成人身伤害。

另一种供应链安全风险在安全行业更常提到,即产品开发的供应链安全风险。如今,无论开发一种硬件还是软件,作为新一代的信息化产品,它必定用到了大量的第三方开源代码。一旦这些开源产品存在安全漏洞,软件供应链就可能受到攻击,造成巨大的损失。2021 年 12 月 10 日,Apache Java 模块 Log4j 库

第一个远程代码执行漏洞被公开,在全世界都导致了广泛的攻击。①

根据 2022 年 12 月 Make UK 的一份报告,42%的英国制造商在过去 12 个月内成为网络犯罪的受害者,其中四分之一遭受了 5 万—25 万英镑的损失。根据 Security Intelligence 的数据,制造公司也成为新近流行的勒索软件最大的攻击目标,占攻击总量的四分之一,居于专业服务(17%)和政府组织(13%)之前。

当然,针对这些风险,制造业也有不少安全科技工具可以使用。比如,区块链技术可以提供分布式和不可篡改的数据存储,可用于确保供应链的透明性、防止数据篡改和验证设备的身份和可信性;AI 技术可以基于历史数据和模式来预测与防范潜在的安全漏洞和攻击,从而快速发现异常和威胁;物联网安全解决方案可以提供设备身份验证、数据加密、访问控制和漏洞管理等服务,以保护连接的设备和传输的数据;边缘计算技术能将数据处理和分析推向接近数据源的设备,从而降低数据被攻击和泄露的风险。

“工业 4.0”也并不限于制造业,事实上医疗、交通等与民生深度相关的行业,在智能化道路上面临着同样的问题。

以医疗行业为例,作为数字化程度相对较高的行业,医疗行业已经采用了大量数字技术,比如:AI 被应用于医疗图像分析、诊断辅助、疾病预测和药物研发;大数据分析被用于处理大量的患者数据,发现潜在的病例模式和治疗趋势;传感器和物联网在监测患者的健康状况、实现远程健康监护和个性化医疗服务中日渐普及;云计算帮助医疗机构存储和管理大量的医疗数据,简化数据管理流程,并提供远程访问和协作功能。

与此同时,根据中国医院协会信息管理专业委员会(CHIMA)发布的《2018—2019 年度中国医院信息化调查报告》,参与调查的 839 家医院中,仅有 43.95%通过了等级保护测评,其中三级以下医院中 75%未开展过等级保护测评。《2019 健康医疗行业网络安全观测报告》数据显示,在对 15 339 家医疗行业相关单位进行观测后,共计发现 1029 家网络服务器存在僵尸、木马或蠕虫等恶意程序的单位,6446 家单位的应用服务端口暴露在公共互联网中,4546 家单位

① 本文关于供应链安全风险的论述,来自科技部“云计算+大数据”国家重点专项课题负责人、数秦科技执行总裁兼首席科学家崔伟在接受本报告前期调研时提供的宝贵意见。

的网站存在被篡改的安全隐患,其中 261 家单位已发生网站被篡改情况。

这种状况意味着什么?2019 年,德国一家专注漏洞分析和管理的公司发现,一些服务器含有大量医疗放射图像,但它们被暴露在了公共互联网上,其中有 14 个服务器系统与中国有关,总计包含近 28 万条医疗数据。这些数据详细记录了患者的个人信息和医疗状况,攻击者借此在暗网上进行交易,获得了巨额利润。

世界已经来到了数字社会的新阶段。许多人认为数字技术是一种特色,有的行业有,有的行业没有。这种想法在 10 年前甚至 5 年前都可能是事实,当时数字技术和其他所有技术之间的界限更明显;但今天,每个行业都在走向数字化,从而提高了技术水平。交通、医疗、制造、建筑、发电等行业,在使用和依赖数字化技术方面相互联系在了一起。

为什么新兴技术要采用数字技术?或者说,为什么数字化是科技发展的内在逻辑?现代科学源于 17 世纪的科学革命,这场革命的核心特征就是将自然科学数学化。唯有如此,我们才能更精确地进行观测、计算、实验与控制。因此,数字化技术也就成了各门科学的核心工具。

以合成生物学为例,它通常是指对现有蛋白质或生物材料进行重新编程,以实现新功能。云计算和 AI 技术的兴起大大推动了这一学科的进展,生物学家可以借此处理更大的数据集,以更快的速度和更大的规模进行基因测序,从而精密地操纵和重新设计这些细胞,来驱动特定的结果、研发新的生物能源。

既然数字技术无所不在,那么数字安全风险自然也就无所不在。只要有网络、有数字信息的地方,就要有安全科技。小到微信登录,大到物流供应链系统的防火墙、疫苗生产系统的设计,都离不开安全科技。只要不出事故,我们常常感知不到这些科技的存在,但它们已经遍布生产生活的方方面面。

因此我们提出,未来有两项通用技术,一是人工智能技术,二是安全科技。技术是把双刃剑,在它以更快的速度变得更加强大时,我们不可能扔掉利刃,放弃科技——唯有以另一种科技来打造足够安全的剑鞘。

三、安全科技的关键技术能力及双重价值

（一）安全科技的内涵及关键技术能力

什么是安全科技？安全科技包括一系列旨在保护信息、网络和计算机系统免受未经授权的访问、攻击和威胁的工具、技术和系统，其市场包括硬件、软件和服务，随着 IT 基础架构向云端的演进，软件市场的占比在不断提升。

在相当长的一段时间里，网络安全（network security）即安全科技。过去 20 多年的时间里，网络安全得到了极大的发展。然而，随着数字化的深入，越来越多的细分领域涌现出来，传统的“网络安全”概念难以进行全面覆盖，而“安全科技”这个词在这样一个转型期被提出来，用以关涉多个领域的技术和理念。它的边界是不稳定的，因为新的安全风险层出不穷，安全技术一直在更新升级。

本质上，安全科技是一种伴生技术。它永远在面向新科技，面向新发展。不管哪种新技术，在推进过程中都会产生新的安全问题。新技术的发展有时候非常快，所以安全技术的发展和创新也同样在高速变动。[①]

从定义上讲，安全科技不仅超出了 network security 的范畴，也超出了国外常用的 cyber security 范畴。因为 security 更多指向与对抗相关的安全内容，而安全科技还需要包含 safety、reliability、transparency 这几个相关但不同的概念，是一个极其复杂的学科。

安全科技包含系统安全、网络安全、数据安全、业务安全、AI 安全等范畴。但在现实中，每一个技术域、每一个业务域都有自己的安全问题，都有自己的垂直分类，每一个垂直领域都很重要，区别主要在于行业侧重点不同。例如，用于保护各种应用（桌面、Web、移动等）生命周期安全的应用安全，保护各类新型终端（IOT 设备、OT 系统）安全的终端安全，保护云基础设施以及用户基于云构建的应用与数据的云安全，还有芯片安全、身份管理、威胁检测和响应、统一威胁管

① 感谢浙江工业大学网络空间安全研究院院长宣琦接受调研时提供的宝贵意见。

理、风险与合规服务等等。

目前,全球行业共有数百家公司分布在各个垂直领域,是一个较分散的市场,且更新换代速度极快,整个安全科技产业的赛道细分远超其他行业。美国拥有 80%—90%的网络安全公司,以色列、英国、法国、德国也正在培育新的明星企业。全球领先的公司有思科、微软、黑莓、Accenture、Akamai Technologies、BAE Systems 等企业,我国的深信服、奇安信、启明星辰、安恒信息、蚂蚁集团、华为、腾讯等企业在垂直领域拥有较突出的技术优势。全球权威知识产权机构 IPR-daily 发布的《安全科技专利分析报告》显示,截至 2023 年 4 月,中国是全球安全科技发明专利最主要布局的国家,共拥有专利申请数 20 445 件,为第二名美国的 2.23 倍。科技型骨干企业成为安全科技专利的主力军,蚂蚁集团、华为、腾讯、国家电网、中兴、中国工商银行、中国移动均位居全球安全科技专利申请前十名。

目前,安全科技的市场仍由传统企业占据主流,但基于云的新一代企业正在崛起。在全新的技术形势和安全需求下,新的参与者很有可能重塑这个分散的市场。

从时间上划分,安全科技的发展大致可以分成三个阶段:

第一,本地保护阶段(20 世纪 70 年代至 21 世纪第一个十年初期)。这一时期的安全科技主要基于密码学和防火墙技术,其保护的对象是网络相关的基础设施。进入 20 世纪 70 年代后,随着个人计算机的普及,各行各业都迅速采用计算机处理各种业务。计算机网络尚未大规模普及时,其主要面临的威胁是来自非授权用户对计算资源的非法使用、对信息的修改和破坏。这时,计算机安全的主要目的是,采取措施以确保信息系统资产的保密、完整和可用,比较典型的是通过操作系统的访问控制手段来阻止非授权用户的访问。

计算机网络,尤其是互联网的出现是信息技术发展中的一个里程碑。1994 年,中国科学院高能物理研究所的 IHEPNET 与国际互联网络连通,这条带宽 64K 的国际专线完成了中国全功能 IP 连接,标志着中国迈进了互联网大门。从 20 世纪 90 年代到 21 世纪第一个十年中期,我国信息化和网络迅速发展,使得围绕计算机的安全正式进入网络时代,信息系统安全(也被称为“网络安全”)产

业实现了从 0 到 1 的发展。

这一时期安全科技的作用，主要是保护信息在存储、处理和传输过程中免受非授权的访问，防止授权用户的拒绝服务，同时检测、记录和对抗此类威胁——病毒、木马、蠕虫，由此产生了对防火墙、防病毒、PKI、VPN 等安全工具的需求。20 世纪 90 年代，NASA 的研究员发明了世界上第一个防火墙。

这一阶段的另一个标志是发布了《信息技术安全性评估通用准则》，即通常所说的通用准则（common criteria，CC），后转变为国际标准（ISO/IEC 15408），我国等同采纳此国际标准为国家标准（GB/T 18336）。

第二，网络保护阶段（21 世纪第一个十年中期至第二个十年中期）。越来越多的设备接入并融入互联网，将传统的虚拟世界与物理世界互相连接，形成了一个新的空间。互联网成为个人生活、组织机构甚至国家运行不可或缺的一部分。此时，以入侵检测、反病毒和漏洞扫描等技术为代表的网络安全是主流。

第三，智能保护阶段（21 世纪第二个十年中期至今）。大数据、AI、云计算等新技术的出现和应用，改变了安全科技的面貌。这个时期的信息安全威胁来源从个人上升到犯罪组织甚至国家力量，“信息安全保障”的概念形成并成熟。智能互联设备成为新的攻击目标，系统安全风险加大。安全科技开始向智能化、自适应、自愈合等方向发展，并从技术扩展到管理，从静态扩展到动态，通过各种安全技术和安全管理措施的综合融合，构建深度的、主动的防御体系。

随着 IT 架构的演进，安全科技同样走向了无边界。在初期和发展期，政府和企业数据基本存储在私有数据中心，员工也基本在公司内办公，几乎所有的资产都在防火墙技术的保护下；到了创新阶段，数据存储转向云端和本地的混合形式，移动化办公也开始普及，办公网络的边界变得模糊；未来，大部分数据将会上云，安全科技也将从被物理硬件塑造转向被软件定义。

经过长期的发展，安全科技领域目前已经有了大量的技术储备。正如前文提到的，安全科技覆盖面极广，不同技术侧重于不同的垂直领域，且都具有各自应用场景下的重要性。根据技术的成熟度与前沿性，我们从应用价值角度将安全科技系统分为四大板块，分别是基础安全、业务安全、AI 安全、未来安全。

其中，基础安全包含传统的物理安全、系统安全、网络安全、数据安全、身份

与访问安全、密码学等，是数字世界网络中任何企业与机构都必须具备的安全防御技术能力；基础安全包括大量已经成熟应用的技术，如密码学与加密技术、防火墙、身份认证和访问控制、公钥基础设施（PKI）、入侵检测系统（IDS）和入侵防御系统（IPS）、漏洞管理和漏洞扫描、威胁情报和情报共享等技术。

业务安全是在基础安全技术保障了基础运行后，企业根据自身业务需求，针对可能遇到的业务风险而研发出的一系列技术。业务风险主要指不法分子利用业务规则漏洞和技术手段，进行数据爬取、账户盗用、交易欺诈等。因此，风控技术是业务安全技术的核心，包括风险识别、风险预警、风险管控、风险处置。近几年，风控技术加速智能化，风控模型从传统的基于规则转变为基于大数据与 AI，一些企业已开始将图技术、隐私计算等前沿技术整合进风控模型。

AI 安全是一个新领域。过去几年，通过应用 AI 来提高安全技术的效率和成功率，已经成为技术领先企业的常态。譬如通过从已有数据中学习，AI 可以更快地识别攻击的模式和趋势，从而预测未来的攻击，并配置自动响应威胁，在更短的时间内对抗网络威胁。随着 AI 大模型的诞生，AI 技术的影响力迅速扩大，不再只是助力安全，自身也成为安全保护的对象，逐渐形成全新的 AI 安全研究领域。目前，AI 安全还处于相对基础的探索阶段，主要是具备 AI 能力的企业为保障自身 AI 技术应用而进行的安全研发，如 AI 算法模型的安全、AI 原始数据的去毒清洗等。

在 AI 之外，面向未来的安全形势，另一些前沿技术也已经出现。尽管目前它们的应用规模很小，但有着相当大的潜力来改变整个安全版图，如量子信息安全技术。

整体来讲，安全科技系统具有许多尖端和前沿科技，包括：

1. AI 和机器学习。通过从已有数据中学习，AI 可以更快地识别攻击的模式和趋势，从而预测未来的攻击，并配置自动响应威胁，在更短的时间内对抗网络威胁。

2. 区块链技术。区块链提供了去中心化、安全性强的数据存储和传输机制，可用于数据完整性验证、身份认证和分布式防御。

3. 零信任安全模型。零信任安全模型意味着“从来不信任，始终在验证”

(never trust, always verify),它假设防火墙两侧的所有内容都可能有害,无论它源自何处,都要求对所有用户和设备进行严格的身份验证和授权,并实时监测和评估其访问行为。

4. 隐私计算。隐私计算是一个包含安全多方计算、同态加密、差分隐私、零知识证明、联邦学习以及可信执行环境等主流技术子项的相关技术体系,可以让数据以“可用不可见”的方式进行安全流通。

5. 物联网安全。随着物联网设备的快速增加,物联网安全成为关注的焦点,包括设备认证、数据加密、远程管理和漏洞修复等方面的技术。

6. 云安全。随着云计算的普及,云安全技术变得至关重要,包括数据加密、访问控制、虚拟化安全和云供应链安全等方面的技术。

7. 生物特征识别和多因素身份认证。生物特征识别技术(如指纹识别、虹膜扫描和声纹识别)以及多因素身份认证技术(如使用密码、指纹和面部识别),可以提供更强大的身份验证和访问控制机制。

8. 量子安全。量子安全是在未来的量子计算时代,传统加密算法受量子计算攻击而失效后,保护信息安全的一种技术。常见的量子安全技术包括量子密钥分发、量子随机数生成、量子认证、量子安全通信协议等。量子计算技术极大地提升了算力上限,并动摇了密码学的基础,目前在全球范围内引起了创新热潮,多个国家和众多行业已经投入对量子计算和量子安全的应用探索中。

(二)安全科技发展的五大趋势:智能化、零化、匿名化、弹性化和量子化

找到安全科技的共性并不难,大多数情况下安全科技发展的趋势是由最新的漏洞驱动的。因此,安全科技的最大特点就是攻防。网络犯罪者无休止地挖掘和利用新漏洞来发起攻击,安全战线一方则尝试消极或积极的防御。

有体育竞技经验的人都知道,一味防守不仅效率低下,而且表现常常难以令人满意。为了提高防御效果,需要预先对攻击者做出判断,比如:未来主要的数字安全风险会出现在哪里?哪些技术有可能成为关键?

近几年,面对新的安全威胁和新的应用需求,涌现出一大批新的安全技术,信息安全专家、中国科学院院士冯登国将这些新技术的特征概括为智能化、零

化、匿名化、弹性化、量子化。这五个特征，也可以概括为安全科技系统的整体发展趋势。

1. 智能化特征。在大模型潮流下，AI 安全必将成为主流的安全关切，该领域主要包括 AI 模型自身安全、应用导致的安全风险，以及 AI 在安全领域的应用等方面。

AI 在安全领域的应用包括防御和攻击两个方面。在防御方面，AI 赋能防御技术提升防御的能力和水平：AI 可有效提高威胁检测与响应能力；AI 可提供较高的预防率和较低的误报率；AI 可准确、快速地预防、检测和阻止网络威胁，识别并分析未知文件；AI 可克服人性的弱点，抵御以人为突破口的社会工程学攻击。在攻击方面，AI 赋能攻击技术提升攻击的精准性、效率和成功率：深度学习赋能恶意代码生成可提升其免杀和生存能力，攻击者利用深度学习模型可提升识别和打击攻击目标的精准性；AI 赋能僵尸网络攻击可提升其规模化和自主化能力；AI 赋能漏洞挖掘过程可提升其自动化水平；AI 可实现智能化和自动化的网络渗透能力。

2. 零化特征。由于云计算和虚拟化等技术的发展，网络边界越来越模糊，零信任安全架构就是基于这样的现状提出的。零信任（zero trust）不是一种技术，而是一整个框架和思想，核心是“从来不信任，始终在验证”。它的工作原理是假设每一个链接都是威胁，无论是来自外部还是内部。零安全技术主要包括零信任架构、零知识证明、零中心技术、零存在模型、零密钥协议等。

3. 匿名化特征。匿名化针对的是未来将越发凸显的数据安全问题，核心是隐私保护技术体系。部分隐私保护技术可用于解决使用中的数据安全问题，这类技术也被称为“数据使用安全技术”，如机密计算（confidential computing）、联邦学习、同态加密、安全多方计算。其中，机密计算是当前最热门的数据使用安全技术，它可提供硬件级的系统隔离，保障数据安全，为多方信息流通过程中数据的“可用不可见”提供安全解决方案。

4. 弹性化特征。弹性的目的是使系统具有预防、抵御网络攻击的能力，以及在遭受网络攻击等不利情况时仍能继续运行的能力，因此弹性是贯穿基础安全、业务安全等领域的概念。弹性安全技术主要包括弹性公钥基础设施

（PKI）、定制可信空间（TTS）、移动目标防御（MTD）、棘轮安全机制、沙箱隔离、拟态防御（MD）、可信计算等，其中弹性 PKI 被认为是新一代数字认证的基础设施。

5. 量子化特征。安全技术通常与计算能力有关，量子计算技术可使计算能力大幅度提升，从而解决现实世界中的复杂计算问题，量子计算技术将直接动摇现有安全技术的基础，因此抵抗量子计算攻击的安全设计理论、安全分析评估方法、安全解决方案及新型困难问题的寻找和优化实现等都成为当前的研究热点。

一方面，如果产学研各界能够把握这些新特征，尽快进行面向未来的安全科技研发，将有效提高数字社会的安全水平；另一方面，我们需要时刻记住，数字安全风险是和数字化转型紧紧相连的，掌握了最先进安全科技的组织和国家也将在产业发展上获得先机，从而为经济发展提供动力。

（三）安全科技的双重价值在放大

总结起来，安全科技有压舱石与助燃剂双重价值：

第一，守住技术的安全底线，防御外部风险隐患，让技术“难作恶”。例如，利用安全科技打击在线洗钱、防御 DDos 攻击、拦截非法访问等。这是数字社会稳步运行的压舱石，也是大部分人对安全科技的既有理解和定位。

第二，提高技术的安全上限，降低技术运行成本，让产业“跑起来”，让更多人受益。这一作用被许多人所忽视，是导致安全科技长期被视作成本而非动力的根源。在新技术的前期发展中，常常出现因为安全隐患而无法投入实践的问题，就好像在刹车系统被发明出来前汽车行业不可能出现。可以说，安全技术是新兴技术真正使得产业起飞的必要助推器。

当前 AI 技术蓬勃发展，适逢其时地为人类有效求解复杂问题提供了强有力的新手段，但随之而来的数据隐私、技术滥用、失控等安全问题引发了全球担忧。加强对 AI 这一新兴技术的潜在风险研判和防范，确保 AI 安全、可靠、可控，已成为 AI 产业发展的核心要素。

四、未来:安全科技将成为新型公共品

(一)“慢起飞模型”:安全科技是技术向善的集中体现

尽管我们在论述安全科技的价值时,聚焦于它对新兴科技的护航、控制以及对产业发展的促进,但必须强调的是,安全科技并不是一个纯商业范围的议题。本报告希望从两个方面厘清安全科技的公共品特征:

第一,产品运作层面。防病毒软件、入侵防御系统等商业安全系统是私有产品,但威胁情报、漏洞信息共享、隐私计算、基础设施保护等技术和平台都具有明显的公共品属性。

第二,价值观层面。长期以来,安全科技属于商业和技术界的话题,关于它的讨论一直与企业保护、IT 基础设施建设等词汇联系在一起,个人和社会在其中缺乏足够的表达。但我们认为,安全科技的完整内涵不只是对攻击者的防御体系,还包括保护平等发展、社会团结与个人尊严的技术与服务。

厘清安全科技背后的价值观,有助于我们加强对安全科技乃至整个科技价值的理解和发展路径的思考。普通人为何要关注安全科技?因为安全科技事关每个人的尊严、人际交往、健康及社会认知。

举例而言,当前相当普遍的电信诈骗,其中的个人信息从何而来?可能是犯罪分子通过非法手段购买得来的,但肯定有很大比例来自受到网络攻击的企业内部数据。尽管企业或许没有从攻击中受到直接的损害,但每个个人都承担了相应的风险。电信诈骗的主要受害人群不仅遭遇了经济上的打击,在心理上往往也承受着压力,部分人群如老年人会因此而被子女劝阻使用数字化工具,这对于人格尊严无疑是一种伤害。更进一步,即使一个人没有受到电信诈骗,他依然可能因为数据泄露而遭受网络暴力,在心理上遭受严重损害。这种心理上、人格上的伤害相当普遍却未被统计,未被纳入安全科技的充分讨论中。

相关学术研究发现,不同人群在被网络欺诈或被在线欺凌后,选择退出使用数字化设备并导致健康状况恶化的比例不同,最容易受到影响的是儿童、青少

年、残疾人、老年人等相对边缘化的群体。这说明，不安全的数字环境将扩大群体间的不平等。

平等发展角度的另一个例子是，大公司往往有更强的安全科技能力，中小型公司也认为部署安全科技能力是顶级规模的企业需要关心的议题。但现实是，勒索软件等恶意软件对中小型企业和中型市场公司会构成生存威胁，而对大型企业通常不会，因为较小的组织在遭遇打击后更可能面临重大的、公开的服务中断。例如，2019 年 5 月美国得克萨斯州一家中型钢结构制造商遭受了勒索软件的攻击，其工具和财务会计软件被永久加密，在难以支付赎金的情况下被迫破产。此外，一旦公司不得不公开处理安全问题，客户对公司的信任可能很难恢复。根据麦肯锡关于数字信任重要性的研究，近 10%的受访者表示在得知数据泄露后停止了与供应商的业务往来。

新技术会造成新的社会问题，如假新闻问题。2022 年以来的 AIGC 热潮加大了该问题的严重性。为了保护信息的真实与完整性，各国政府正在制定新的法律法规，发展新的标签系统和水印技术等，帮助人们识别虚假内容。可以预料，在 deepfake 技术的蔓延下，假新闻将比算法导致的信息茧房更进一步造成社会撕裂，加大社会成员间的沟通成本。

技术不仅是工具的集合，也反映了我们的价值观，安全科技是其中的典型。当我们讨论投资安全技术时，我们讨论的不仅是代码与数字，也是科技的伦理、社会的结构、人群的互动和个人的安全感。

为此，我们认为数字化转型应该采取“慢起飞模型”，在鼓励新技术发展、产业升级的同时，不仅要追求转型的速度，更要在社会需求方面投入更多力量，通过更好的安全科技和制度构建，尽量减少个人与社会为数字化转型所付出的代价，让起飞能惠及更多人。①

“慢起飞模型”与技术的创新冲动、商业的盈利需求之间一定程度上存在矛盾，要想切实实现“慢起飞”，需要多方的努力。一方面，要求企业真正以 ESG 的思维进行运作，不仅要投入更多资金和人力加强安全科技能力，并且要转变理

① 感谢中国社会科学院大学哲学院教授、中国社会科学院科学技术和社会研究中心主任段伟文在接受本报告前期调研时提供的宝贵建议。

念,不再仅将安全视作企业运行的成本,而是当作企业的社会责任和社会影响力的核心落脚点,是技术向善的重点领域;另一方面,这也要求安全科技真正成为公共品,由国家、社会和产业界共同推动新安全技术的低门槛、规模化应用,建立数字化安全风险公共治理空间。

(二)政企合作是安全科技成为公共品的核心驱动力

那么,如何让安全科技成为公共品?我们认为,推动力可能有四种:合规激励、商业保险、标准建设、安全思维。

一是合规激励。数据爆炸正催生越来越多技术、数据链路方面的复杂性难题,然而大多数企业往往更倾向于把资源投入 IT 研发中而不是安全保障上。目前,国内大多数企业和机构在 100 名 IT 研发人员中,配备的安全工程师不到 0.5 名。在大部分企业自身没有直接动力来投入安全防御时,政府则需要推动相关的保障投入。在这方面,欧盟 2016 年通过的《通用数据保护条例》,对于产业界起到了重大的指导意义。我国近些年在数字化安全领域同样颁布了《网络安全法》《密码法》《数据安全法》等法律,为企业建立起了基本的安全框架和安全心态。但其中多为定性规定,数字安全事件惩罚的定量及赔偿标准依然难以确定。下一阶段,合规的细化和惩罚制度的逐步实践将成为安全风险改善的重大推力。

二是商业保险。商业领域的创新是落实合规的重要助手,商业网络安全保险也能有效帮助建立数字安全事件评估及赔偿标准。网络安全保险由私人保险公司向企业和其他实体提供,保护第一方(投保人)和第三方(投保人客户)免受网络事故带来的损失。网络安全保险有两种形态,一是只覆盖网络风险的独立保单,二是作为一般保险(覆盖多种类型风险)的一揽子保险范围的一部分(如一般商业责任保险)。随着该险种的发展,市场接受率也在不断提升。

能够为合规买单的企业基本只有大型企业,从商业运作角度来说则是保险公司。由于保险公司运作要求对风险有精确的评估、监控和响应,因此商业网络安全保险市场的发展将显著提高这一领域的透明度,将安全风险从企业内部问题变成更公开的事务。另外,不同企业的保费评估能有效补充法律的定性规范,在不作为和直接触犯法律之间设置精细的处罚梯度,引导企业与机构改善自身

的安全表现。因此，商业网络安全保险有利于政府与企业、市场合作，共同攻克业态尚未成熟时期的发展瓶颈。

三是标准建设。标准建设同样是一个生态合作范畴，借助政府、头部机构与企业的带头作用，为全社会的安全技术建设提供标准。标准建设分为两个部分。第一，具体产业中的标准建设，如上文提到的 IIFAA 联盟对车联网数字钥匙的技术规范等。没有清晰的标准就会导致责任认定不清，特别是在新技术落地方面，大大增加了参与方的顾虑。第二，更高层面的安全框架规范，如美国国家标准与技术研究所制定的网络安全框架，以美国政府运行多年的《联邦信息安全管理法》经验集中输出为标准，旨在“确定现有的自愿共识标准和行业最佳实践，以将其构建成网络安全框架”，以最佳实践方式推动安全。由于参与制定安全框架标准的专家通常是中小公司无法雇用的顶级专家，因此充分利用好安全框架准则可以有效帮助中小型组织实施高质量的安全措施。

四是安全思维。我们将重点讨论两种思维转变的需求：以人为本的安全设计、运维与建设并重的安全构建。根据世界经济论坛的《2022 年全球风险报告》，95%的网络安全漏洞是由人为失误造成的。显然，个人在对抗网络风险方面承担了不相匹配的责任。一个企业乃至一个社会基础设施的网络安全不应依赖个人的持续警惕，必须要求有更多专业资源的机构通过更有效的防护机制建设来减少人为失误的风险。以人为本的安全设计就是其中一个必要的机制，它要求在整个安全风险的控制管理流程中，优先考虑参与者的体验，而不仅仅是考虑技术。运维与建设并重的安全构建，针对的是当前我国在安全方面的一个普遍倾向，即企业和政府部门构建信息化系统时，普遍存在“重建设轻运营”现象。这意味着企业在进行信息系统建设和开发时，往往只关注硬件设施建设，而忽略了持续更新维护系统的重要性。然而，数字化的特点之一就是从量变转向质变的速度很快。信息化系统运维必须有足够的运维量，这要求我们的安全构建思维从传统的硬件模式转向更偏重软件的服务模式。

（三）结语

正如数字化安全风险以及安全科技的内涵所体现出的生态融合特征，合规

激励政策的细化、商业保险市场的发展、行业标准的建设、安全思维的更新，都要求政府、企业、学术界以及公众各个层面的合作。安全科技作为一种伴生技术，本身便具有浓厚的伦理取向，需要社会各界的积极关注与参与。这种关注与参与，是我们在数字社会中安全、平等、团结与有尊严生活的必修课。安全科技作为未来的通用技术之一，尽早对它进行投资，支持我国企业在相关领域的研发和应用，将有助于掌握新兴技术发展与产业迭代的主动权，使安全科技成为建设世界科技强国的基石之一。