《美国数学学会通报》（新系列）第 00 卷，第 0 期，
第 000–000 页 S 0273-0979(XX)0000-0

arXiv:2305.02329v1 [cs.LO] 3 May 2023

机器时代的证明

安德鲁·格兰维尔

Abstract. 我们比较了与（传统的）基于社群的证明验证相关的价值，以及与计算机证明验证相关的价值。我们提出了计算机证明可以借鉴人类经验中成功策略的方式。

1. Introduction

专业的纯数学家也助长了本科生所抱持的幻想：数学稳固地建立在公理基石之上，形成一座由客观证明为真的结果构成、不可更改的结构。[1] 然而，任何有经验的研究者在实际工作中采取的方式都大不相同。他们以一个知识库为基础（其中既有论文、预印本和书籍，也有默会知识），并假定且希望这些知识是正确的（或者至少可以修正）。在实践中，形式证明——将证明一路追溯到公理——既乏味又几乎毫无意义，因为进展主要以新增并纳入知识的内容来衡量，着眼于遥远的地平线，而不是对早已远离我们的内容进行繁琐且难以跟进的严谨审查和安全检查。另一方面，直觉证明不仅能为研究者群体所接受的论断提供合理依据（至少目前如此），还能激发、推动并启迪新一代的发展。

然而，变化正在发生。随着实用的证明验证系统问世，这些系统配有大量现成的定义库，以及从公理中必然推演出的丰富基础定理，研究者或许可以输入自己惯用的直观证明（尽管需要作出修改，以适应证明系统已知和所需的内容），从而获得形式验证的证明，并更加确信自己没有遗漏什么简单的问题。

此外，一旦机器能够标准化地将直观证明转换为形式证明，反之亦然，那么机器当然也能协助证明过程，起初可以确认所提出的引理（通过识别其库中已有的证明，或找出对现有证明的直接修改）。在某个阶段，我们或许会期待此类机器证明验证器被用作“黑箱”，届时我们无需理解它们正在做什么，也能相信它们找到了客观的证明。纯数学研究的很大一部分都集中于证明所提出的命题，而我们很难猜测机器参与这一过程是否会存在任何限制。这引发了许多关于数学研究未来以及人类在那个未来中的角色的问题，正如.....所强调的

2020 *Mathematics Subject Classification.* 主要。[1]一个美妙的巧思，让一代又一代的本科生相信数学是美丽且永恒的。

©0000（版权所有者）

Venkatesh 的文章 [17]所强调的那样。这里我将聚焦于一个重要方面：对于机器辅助形式证明，我们是否会比对当前较为直观的证明更有信心，认为它已得到客观证明？如果会，依据又是什么？要恰当地思考这些问题，我们或许首先应当了解当前验证证明的体系及其优势（关于其中许多问题的更多细节，见[8]）。我们还应当更好地理解机器证明验证器究竟在做什么，包括负责将证明输入机器的人类团队所发挥的作用。

证明简史。数学家通过将新的命题归约为已被接受为真的命题，来证明新的命题：

> 如果……理解正如我们所提出的，那么证明性的理解就必须依赖于那些真实、基本、直接，且比结论更熟悉、更先于结论并能解释结论的事物……—— Aristotle

这个递归过程必须有某些起点，也就是一组公理；理想情况下，这些公理应表述优雅、数量尽可能少且彼此一致（我们不应该能从公理中证明一个命题及其否定！），以此开启证明过程。此外，公理最好能让我们证明所有感兴趣的定理；如果做不到，我们也应始终能够提出新的公理来提供帮助。

据我们所知，欧几里得最早认真尝试提出公理，而这些公理经过完善和发展，一直延续到20世纪。康托尔尤其指出，数学进步取决于概念创新，因此我们应始终准备好调整公理，而这些公理应显而易见地为真且彼此一致。然而，弗雷格广受推崇的体系并不一致（罗素在对其语言的解释中发现了悖论）；随后，罗素和怀特海德提出的替代体系也并非完全“不言自明”，而且难以实施。

这些问题促使希尔伯特提出，我们应该能够从任意一组相容公理出发（配以定义明确的简单语言，以及适用于证明的恰当推理规则），看看由此会得出什么；换句话说，我们不必从永恒不变的真理出发。是证明将数学创生出来。弗雷格对此并不同意，他认为，线就是线，也就是日常语言中所说的那个物理实体；如果你提出一种理论，凭借一组晦涩却相容的公理得出了不同的东西，那么就应该摒弃这套理论。

希尔伯特的规则不可避免地会导致公理系统的多元化：例如，整数所对应的无穷与实数所对应的无穷之间，在大小上是否还存在一个无穷，这一问题独立于 10 条“标准公理”（ZFC），正如戈德尔和科恩所证明的那样。允许采用这类替代性基础，似乎会把数学变成一场遵循特定规则的游戏，而不是建立在深刻洞见之上的发展；但如今的数学家正是运用康托尔的规则：我们只接受那些能够带来有趣创新的可能性。

希尔伯特的希望是找到一组相容且完备的公理（能够让我们证明所有定理），并且这一点还可以得到证明。然而，1931年，库尔特·G¨odel 用不完备定理从根基上动摇了希尔伯特提出的数学。这些定理本质上表明（粗略简化地说）：不存在一组相容的有限公理与规则集，能够用来证明关于整数的所有定理，也无法证明自身是相容的。灾难。G¨odel 的定理直接意味着

> 经典数学不可能得到严格的论证。—— John von Neumann

危机应对。大多数纯数学家通过忽视这场无法解决的基础危机来应对它。[2] 对于许多数学家而言，G¨odel 的质疑似乎与他们正在研究的内容无关。他们需要一个在所有“合理情形”下都可靠的形式推理系统（正如近年来范畴论的重新兴起所体现的那样）。在过去一个世纪里，基本公理系统（ZFC）一直被接受为大多数现代纯数学本质上未变的基础，尽管有 G¨odel 的结果，而且它确实有效。但没有任何办法绕过 G¨odel 的定理：一些作者错误地认为，计算机不知何故会有所帮助，但这源于对丘奇–图灵论题的不了解。这一工作假设认为，所有合理的计算系统都是“等价的”且“通用的”。这意味着它们各自都能计算一切可计算的事物。因此，它们可以相互模仿，而计算机能够完美地模仿人类以及人与人之间的互动。[3] 此外，人类也能够完美地模仿计算机，因此我们的根本局限是相同的。

> 数字计算机……旨在执行人类能够完成的任何操作。——
> Alan Turing

还有一场不那么为人所注意的基础危机。简单的计数论证表明，NP族的可证明正确陈述（即存在关于问题长度的“多项式规模”证明的陈述）在 ZFC 中所有可证明正确陈述族的集合中，只占一个相当小的子集。因此在实践中，没有任何机器、算法、人类或计算机能够指望证明从这个或任何其他合理的公理系统中推导出的所有正确且可证明的数学陈述，甚至连其中极小的一部分也做不到。[4]

2. Accepting a result, by being convinced by a proof

我们愿意相信，纯数学拥有客观证明标准，任何定理都能经过严格论证，最终追溯到公理。但由谁来验证，又如何验证？人们希望有一位技艺高超、客观公正的验证者，一位值得信赖的权威，能够确认语言和演绎规则从头到尾都得到了有效运用。在形式证明中，每一步推理都清楚列出，因此证明可以由机器验证，无需依赖直觉。[5] 但这样的证明能说服数学家吗？

---

[2]我们住在一座极高、看起来结构坚固的塔楼顶层。这座塔楼历经数代建成。工程师告诉我们，地下室发生的火灾已造成无法修复的结构损坏，并建议拆除。然而，他们没有什么好办法能重建出一座高度接近现有水平的塔楼。因此，我们希望等到烟尘和喧闹大多散去，也习惯了持续不断的警报声之后，便能继续若无其事地生活，在上面加盖新的楼层，不去担心下面发生了什么。[3]在模仿游戏中，[16] 审问者试图区分两者：一个试图证明自己是人类的女人，以及一台试图蒙骗审问者、让其相信自己是女人的计算机。（在图灵之前的原版游戏中，扮演计算机角色的是一个淘气的男人。）能够完美模仿人类的计算机就通过了“图灵测试”。[4]还有一个次要问题——是否存在一种算法，能够在多项式时间内为给定 NP族中的定理找到证明？如果存在，那么如果 P=NP，我们试图做的一切都可以完成……不过，我认为大多数人相信 P≠NP。—— Avi Wigderson。

[5]尽管这一理想预设了验证者对每一步的含义和解释达成共识，并且验证者所遵循的逻辑和语言不可侵犯。

> 证明……据称确立了某项数学内容无可否认的真理，[但] 证明实际上并不能让数学家相信这一真理……真正起作用的是别的东西。— Eugenia Cheng [2]

她声称，数学交流通过可信推理，将作者的信念转化为读者相信的真理。这种推理以措辞谨慎的直观证明为结构，并具备足够的演绎严谨性，以消除对歧义或误导的担忧。因此，它符合数学共同体对已知事物及应有之物的感知，这表明验证者应将任何新主张置于更大的背景中，以帮助读者理解。

事实上，对于高度复杂的进展，只要专家能够充分理解整体策略，并且证明的每个部分都能够由精通相应技术的专家进行验证，数学共同体就会接受这些证明。此外，验证过程还会将思想传播到研究共同体中：验证者很少仅仅是为了表示认同而进行验证，他们更希望将新的思想和技术纳入自己的知识储备。他们并不是简单地阅读作者的文字来实现这一点，而是通过用自己的方式重新诠释文本（或讲座内容），并将新思想与自己已知或相信的内容相互联系起来。

相信一个证明：共同体视角。即使只接触到论证的一部分，只要它符合我们对该学科领域的理解，而且留给读者的细节看起来合乎常规，也就是“可靠”，即便有些小错误也应该很容易修正，我们就会相信这个证明。此外，作者对自己作品的信心也能因同事的审阅和提出的好问题而增强：

> 没有哪位……数学家能对自己的学问如此精通，以至于刚发现一个证明，就立刻对它完全确信……他每次重读自己的证明，信心都会增加；而朋友们的认可更能增强他的信心；学术界普遍的赞同与赞誉，则使他的信心达到尽善尽美的程度。—— David Hume（1739）

对相似问题感兴趣、因而熟悉该领域所用技巧的同行，会认可证明是正确的。[6] 为了验证证明，他们会把它放在自己理解的语境中；深厚的知识使他们能够略过并接受大量默会知识，同时核查那些不同于以往做法的（相对少数的）细节。同行希望拓展自己的直觉和范围，而不只是赞同证明中的论证是正确的。读者并非被动接受者。她希望理解、综合这些思想，并将其用于自己的研究。

不同的人从阅读中获得的东西各不相同，因此，一篇新的研究论文可能会沿着此前未曾预见的方向激发新的想法。同一个人在不同时间阅读一篇文章时，也可能获得不同的东西；我们的理解确实会随着时间而变化，有时甚至连我们接近整个领域的方式也会改变。按照这些共同体标准所接受的证明可能是错误的，因为验证者并未仔细检查其中的细节（事实上，大多数细节通常都不那么重要，因为有经验的读者可以自行还原这些细节），但验证者并不是

[6]理想情况下，审稿人应当真正独立，但在实践中，对于基于文献的高水平研究，除了专家之外，没有人有兴趣也有能力审查工作，尤其是当研究足够复杂、需要投入大量时间时。此外，专家知道一项新工作中哪些内容大致属于常规，哪些是新颖的，因此能更好地将精力集中在更可能出现可识别错误的地方。

在寻找严格的证明，而是在寻求一种能够将新工作置于其背景中加以理解的认识。[7] 证明与理解并非同义词，我们必须认识并接受它们如何契合，以及它们有何不同。

如果一段数学内容关乎“理应如此” [2], ，而非“有用、有趣、引人入胜、优美、经过详细证明”，就会让人觉得是对的；正是这种观念促使许多人采用这种方法来证明定理。事实上，保罗·埃尔德什声称，一位客观的至高存在拥有一本《书》，其中收录了每个真定理的完美证明，而且每个证明都简短而优雅。简短，便于验证；优雅，则让人知道该陈述契合得如此之好，必定为真。

审稿工作的一部分是判断投稿内容的价值。Venkatesh [17]认为：“我们赋予一项数学工作的价值完全是主观的，因为它完全取决于我们对该工作的感知，而非任何客观质量。”长期以来，数学的发展一直受到这种主观性的影响。众所周知，Littlewood 曾向审稿人提出三个问题：“它是新的吗？它是正确的吗？它令人惊讶吗？”这些问题至关重要，因为作者的工作发表在哪里，会影响其职位、地位和科研资助。我们能否建立客观标准来评判新的研究成果，还是必须继续坚持主观判断？证明验证系统能否模仿（或超越）人类的判断？

证明的稳健性。我们希望建立一种评审体系，使证明能够保证定理成立，从而使有能力的数学家无需亲自核验自己所使用的每一个证明。熟悉程度能够简化审查过程（尽管这可能在一定程度上助长从众倾向），使论文得到更快速、更准确的评审，并将重点放在根据评审者的经验最有可能引发疑虑的细节上。这种做法支撑了人们对证明稳健性的信念：我们相信，经过广泛使用的技术工具不太可能出现严重问题，因此会对哪些内容需要核验作出相应的假设。即使确实存在错误，经验也表明，只需进行简单的修改，通常就足以使论证成立。任何有经验的研究者在开展自己的研究工作时，都会经常这样做。因此，当他们发现他人工作中存在一些细微的技术性缺陷时，往往倾向于相信这些问题是可以修正的。

形式证明与直观证明的共同点。所有证明的构建方式都大致相同：从双方认可的公理出发，我们构建给定命题的证明。为了取得更大的进展，我们需要避免每次都回到公理，因此需要建立一个命题库，其中存储着我们已知为真的、且表述明确无歧义的命题。传统上，这个命题库保存在研究论文中，并汇编于书籍中。机器也可以使用自己的语言来存储这些内容。

读者希望沿着一条线索追溯到她认为某个论断无需进一步解释的起点。通过交互式链接，读者可以逐步深入探究，无论是形式证明还是直观证明。[8]

乍看之下，这番讨论似乎表明，研究者可以通过对他们从已知结果库中引用的内容进行逻辑推演来取得进展。然而，研究者通常是通过猜想与证明，以及

[7]而且失误可能会发生：在 [8] 中，我们讨论了一些世界顶尖数学研究中的著名错误，这些错误曾一度被学界接受，包括 Voevedsky、Biss、Mochizuki 的工作，以及在较小程度上的 Wiles 的工作；并探讨如何根据公认的证明标准来理解这种接受。[8]帕特里克·马索 [11]最近宣布，软件工具正在开发中，旨在自动将形式证明转换为这类人类可读的交互式证明。

在具体背景中发现新的问题，逐步加深对研究主题的认识；已知结果的精确表述则起着次要作用。

这是截然不同的过程。形式化证明会像一个不知疲倦地追问“为什么？”的孩子一样，将证明的细节一路追溯到公理（直到回溯到不可改变的真理）。但是，在这一过程结束时，孩子还记得自己一开始问了什么，以及自己是如何走到终点的吗？当形式证明验证者报告说一个证明是正确的时，我们是否就应该比那个被告知“因为我这么说了”的孩子拥有更多的信念？

共同体中的证明验证者通常会从工作中学习，不仅能够（原则上）重新构造证明，还能将所学到的思想应用于其他问题。共同体参与者期待的是他们能够理解、阐释、欣赏，甚至在可能的情况下加以运用的证明。他们可能会因找到另一种证明或更清晰的证明而感到兴奋，尽管在形式系统中，这似乎并不起任何作用。

永恒的真理？Venkatesh [17] 指出：“证明是……一种能够促成共识的论证”，而 de Toffoli [5] 则认为，“分享[数学论证] …… [是] 数学论证获得正当性的必要条件”。但是，如果证明能够从公理稳固地推导出最新的定理，我们为什么还需要分享？De Toffoli [4] 认为，“严谨证明的可接受性标准并非一成不变……而是取决于特定时代的数学共同体”，并以基础微积分中的证明为例。这种认为证明以及我们提出的每一个问题都属于最新范式的观点，早已得到承认：

> 一个三角形。这似乎极其简单，你会以为我们……对它了如指掌…
> …即使我们证明了它具备我们所能想到的所有属性，也可能有另一位数学家在未来某个时候（也许是一千年后）发现它还有更多性质；所以我们永远无法确定自己已经掌握了关于这个三角形的一切。——René Descartes （1648年）

如果证明和理解是不变的，这些想法就没什么道理，尤其是在各个步骤都经过客观验证、确保其在任何情境下都成立之后。因此，只有当数学研究的是客观真理时，形式化证明才可以说比直觉证明更有用／更永恒。

3. Myths of scientific objectivity

哥白尼学说曾长期被《圣经》中的段落客观地驳斥。种族主义和性别歧视也曾长期得到所谓客观的社会证据支持和维系。然而，如今这些“客观立场”客观而言都是主观且自利的（关于这一主题的更多内容，请参见 [9] ）。但数学中的客观证明，难道不确实是从表述严谨的公理、语言和演绎规则中推导出来的吗？[9]称某个论断客观为真，便赋予了它权威；这是一种旨在消除疑虑的修辞，但它究竟有何依据？是否存在一种非主观的定义？（也就是说，

[9]事实上，1895年，皮亚诺指出：“不精确的观念无法用符号表示。”他由此推断，只要恰当地使用符号，一切就会变得精确，所得出的真理也将通过客观推演得出。然而，即使我们认同这是一种实现精确的方法，人也可能做到精确，却仍然出错！

不受其他人类诉求影响）如果没有，这类确定性主张真的值得追求吗？[10] 客观性的关键组成部分是什么？

> 科学客观性……表达了一种理念，即科学主张、方法、结果以及科学家自身……不应受到特定视角、价值判断、群体偏见或个人利益的影响。[11]

在任何科学领域中，客观性有多大程度上可以实现？

卡尔·波普尔将“客观性”定义为忠实于事实，即不受预先设定的视角所产生的偏见影响，并能够根据不断变化的信息进行调整。然而，这并不意味着客观性必然能够正确识别永恒的真理。他认为，随着我们对数据的收集和理解不断改进，理论会逐渐接近真理，并随着时间的推移变得更加客观。因此，在波普尔看来，客观真理始终是一个摆在我们面前、值得我们不断追求的目标，但却无法通过客观方式真正实现。

科学实在论认为，客观真理独立于感知而存在，[12]然而，希尔伯特提出的多种可能的公理基础（以及数学著作中给出的不同定义）表明，许多不同的感知都同样有效！即使假设我们从相同的公理和定义出发，每位研究者的语言使用方式和观点也可能导致细微差异，并产生同样有效的解释（这些解释可能相互补充，但要对此做出确定判断并不容易）。无论如何，究竟由谁来决定谁的感知才是客观正确的呢？

库恩 [10] 尤其怀疑客观性的主张，他认为所有观察都是透过主导范式的视角进行的，并由最新（但并非最终）的理论假设加以感知和概念化。[13] 范式引导科学家开展工作，并设定共同体标准。[14] 这种观点让人难以相信存在理论无关的语言、观察，甚至客观性；然而，客观性这一概念在某种特定范式内或许仍有其含义。

有人可能会认为，实验或理论的精确性、简洁性、连贯性和范围是价值中立的，因此应当成为偏好的依据。然而，如果我们写成准确性、优雅性、与理论的契合度以及与理论其他部分的联系，那么这些概念就都不可避免地嵌入了某种范式之中。

过去五十年间，数学领域发生了巨大的范式转变，不同分支受到的重视程度也随之改变，这源于格罗滕迪克的强大影响力。他对代数几何、泛函分析以及许多其他领域的重新思考，使数学界很少关注那些与格罗滕迪克的工作没有密切关联的领域中的杰出成果。例如，组合学被贬低为“不过是计算”（正如我听到阿蒂亚在

[10]“最佳实践”是官僚目前用来以消极攻击的方式羞辱怀疑者的一种说法。“客观真实”会更好吗？ [11] 关于这段引文，以及本节思想更为全面、少些主观色彩的版本，请参阅《斯坦福哲学百科全书》 [14]。[12]无论你觉得热，还是我觉得冷，一个房间都有确定的温度[13]但实验性证据肯定是客观的吧？然而，只有在实验装置可靠的情况下才是如此；而通常只有当它证实了依据我们先前信念得出的结果时，我们才相信它是可靠的。因此，实验性证据只能从其自身范式内部的某种视角来理解。[14]例如，在考察狭义相对论对牛顿力学的改进与取代这一广为流传的例子时，这一点可以与波普尔的观点相一致。

九十年代末所宣称的那样）[15]；而后经过漫长等待和激烈争议，菲尔兹奖才再次颁给狭窄主题范围之外的研究。

无论如何，数学客观性受到语境影响的程度，相较于其他科学受到气候否认、一般政治因素，以及制药和食品公司等大型企业需求所施加的压力，还是要小一些。不过，近来有一些值得注意的例子，例如工业界和政府中的密码学、疫情期间对统计数据的解读，以及如今机器学习行业有时夸大的主张。

定义的困难。对于现代数学研究的许多概念而言，单一的定义并不适用，但所有定义都应彼此协调。[16] 这可能会给形式化验证器带来问题，尤其是在前沿领域，许多相互竞争的方法（包括定义）都可能经过尝试，而只有实践经验才能决定哪种方法在多数情况下效果最好（这些判断也不可避免地带有主观性）。

所有形式语言都必须在如何阐释思想方面作出选择，这意味着有些想法会得到强调，有些则会被弱化，甚至排除在外。[17] 因此，有些新进展很容易用语言表达并加以研究，另一些则不然。要让这一点与客观性的观念相协调，并非易事。

客观证明验证的不确定性原理。数学实践史表明，对一个证明质疑得越少，它就越容易出错。这一重要原理有力地表明，我们必须寻找多种多样的方法来解释和验证任何给定的证明，即使是计算机证明；还应尽可能从多种不同的角度审视它。

4. Computers and proofs

目前，计算机在证明中主要有三种用途：

• 用于建立证明的计算：作者可能会将问题转化为一个包含大量有限情形、无法手工逐一解决的问题，然后借助计算机排除这些情形（例如四色定理（4CT）、关于三维球体堆积的开普勒猜想、魔方的上帝之数，以及有限单群分类的证明）。作者可能需要构造一种具有精细特殊性质的工具，而这种工具可能只存在于高维空间中，只有借助计算机才能切实找到（例如梅纳德在研究素数间的短间隔时使用的400维筛法）。作者也可以利用计算机计算大型实例，从中获得启发，进而加深理解并找到证明；这是《计算数学》论文的传统研究方式。近年来，我们也看到了一些运用机器学习取得的惊人成果（参见，例如 [3][18]）。

•协助核验作者论证的逻辑，或许还可以进行交互式核验，即“计算机辅助证明”：这在核验冗长而详尽的证明时最为有用，因为这类工作连天使都望而却步。例如，一个团队曾与

[15]的确令人遗憾的是，塞梅雷迪在 1975 年关于非稀疏集合中结构的杰出研究，直到 2012 年才获得阿贝尔奖，也因此才受到数学界广泛认可。 ′我们在定义实数或代数和几何中的同义概念时，早已习惯这种做法。 [15] 口语也是如此。会说两种语言的人会发现，有些概念用一种语言解释更自然，有些则用另一种语言解释更自然。[16]不过，也要警惕炒作，例如“用人工智能引导人类直觉”这个标题。这种自我吹捧的说法，既描述了计算机在过去 50 年里于数学领域发挥的主要作用，又将这份功劳据为己有。
1718

Gonthier 合作，使用 Coq v7.3.1 证明助手核验了 4CT 的最新证明（该证明包含 32 个步骤，涉及 633 个需要化简的子图）；另一个团队则与 Hales 合作，使用 HOL Light 和 Isabelle 证明助手核验了他对开普勒猜想的证明[19] 。

Lean 近期取得了令人瞩目的进展，最值得注意的是对 Clausen 和 Scholze 的一项关键结果进行了验证、纠正和改进，本书其他地方对此有所讨论。如今，许多人都在使用 Lean，已被编码其中的思想无论质量还是数量都非同寻常，数学的许多领域也正在见证非平凡结果得到验证。此外，随着更多研究人员为该系统作出贡献，人机交互应会逐渐接近数学家高层次的实践方式。经过进一步改进输入和输出语言，这样的系统便可变得易于使用，并成为数学家工具箱中不可或缺的一部分。

这些证明验证器具有交互性：用户通过交互将证明拆解为机器已知的更简单对象，并向 Lean 提供提示。证明助手会根据当前的库，判断某个陈述是否“显然”为真或为假。如果无法判断，用户就输入更多细节。因此，证明助手会促使用户以严谨的方式解释论证，并补充人类数学家可能觉得无需说明的简单步骤。[20] 例如，对于一个包含十个引理的证明，定理证明器可能很快就能识别并解决其中一些问题；对于另一些问题，则可能需要用户提供更多细节，直到它找到证明的路径。这样一来，程序便能不断学习、维护一个库，而且下次遇到类似问题时或许会更有效率。

证明助手无法判断一个数学陈述是否有趣或重要，只能判断它是否与已证明的内容一致。最终，它们应当（也许很快就会）需要更少的帮助，或许是少得多的帮助。最终生成的证明并不适合人类阅读，但多亏了 Massot 及其团队[11] ，这一点应该很快会改变：他们将机器证明转换为人类能够理解和欣赏的高层论证，从而让我们更加信任其输出。然而，如果证明助手实际上只是通过自身的内部逻辑进行检查（而这种逻辑可能会传播某个细微错误），我们真的能够相信它们是正确的吗？最理想的情况是，这些证明能够得到独立验证，也许由在不同机器上运行的不同程序来完成；实际上，我们提议在计算机证明验证者自身的社群内，对其输出进行审稿！[21] 我们可以基于已经奏效的事物来设计未来。

- 证明声称成立的定理、“计算机生成的证明”：机器学习通常通过巧妙的算法，以简单的方式逐步形成自己的理解。建立大型数据库，并用专门设计的工具对其进行分析，可能会取得惊人的成效（例如谷歌翻译或 ChatGPT），但这与培养直觉（甚至有效地模拟直觉）并不是一回事。[22] 目前还没有“会思考的机器”。

[19]原始证明极其复杂，以至于一个由十二名审稿人组成的团队花了四年时间来验证该证明，最后也只能报告说，他们“有99%的把握”认为证明是正确的。[20]Scholze 和 Gonthier 都表示，在这一繁琐的输入过程中，他们学到了很多东西！[21]这要求他们共享一种共同的语言。[22]围绕“机器学习”以及其他一些形式的“人工智能”，存在大量资金投入和宣传，但许多被大肆炒作的进展，要么被夸大了，要么很容易用设计精良的算法和非凡的计算能力来解释。

人们对创造力和直觉，以及人类如何从一种理解转向截然不同的另一种理解，所知甚少。要让机器模拟这一过程，似乎还遥不可及。[23] 在计算机出现之前，图书管理员常被认为比实际懂得更多（因为他们掌握着大量知识）。计算机是更庞大的知识库，更容易获取，也较少受到他人限制，还能完成一些令人惊讶的事情；因此，人们认为它们具备其实尚不拥有的能力，也就不足为奇了。

计算机拥有巨大的内存，而人类则必须及早运用“策略”来剪枝搜索树，以避免组合爆炸。例如，人类可能会

— 识别陈述中最有可能起关键作用的假设；— 确定一个子目标，将证明分解为更小的步骤；— 识别一种已知的证明技巧，并将其调整应用于当前的新情境；— 利用少量示例猜测相关性质，并快速剪枝搜索树。

所有这些方法都难以在计算机上实现，那么，哪些策略最适合计算机生成的证明？Ganesalingam 和 Gowers [6] 旨在设计一种计算机验证者，使其能够像人类一样学习和思考。他们探讨了应该按照什么顺序尝试不同的策略、计算机能否从过去的经验中学习，以及我们能否设计出一种更好地模拟人类选择的理论，或者将两种方法结合起来使用。[24] 其中一个目标是改进并赢得图灵的“模仿游戏” [16], ，但这一次，机器生成的证明应当与优秀的人类证明难以区分。[25]

计算机错误。如果计算机上的形式证明要具有客观真实性，那么计算机就应该变得毫无错误。人们常常认为，只要计算机程序可靠，它就“不会出错”，但这与他们自身的经历大相径庭。可靠系统投入了大量资源，支撑着你的信用卡、电话、航空公司和大学，而我们都曾为此感到沮丧！除了硬件和软件问题，还可能存在编程问题（例如没有考虑到你的具体情况）或实现问题（例如菜单含糊不清，或说明具有误导性）。难道随着时间推移，这些问题都不能逐一解决，让计算机不再出问题吗？然而，升级或许能处理此前未妥善应对的情况，但也会带来新的功能，从而引发新的问题！而且，如果某个系统最终变得“完美”，我们又该如何证明这一点呢？

计算机硬件故障通常可以通过输出结果来识别：这些结果显然与其他信息不一致。然而，如果计算机正在计算某些我们知之甚少的内容，这种方法就帮不上忙了！制造商很少透露对其产品的担忧（以免吓跑潜在买家），因此即使出现故障，也不太可能告知用户。此外，对于制造商来说，修复广泛使用的计算机芯片中的一个小小数学错误，是否值得付出成本？[26]

[23]想深入了解近期机器学习进展背后的种种因素，可参阅梅兰妮·米切尔的精彩著作《书》 [12] 。此外，也可以看看 Gowers 最近的项目：https://gowers.wordpress.com/2022/04/28/announcing-an-automatic-theorem-proving-project/#more-6531 。[25]在《书》 [6], 中，Ganesalingam 和 Gowers 选定了一些待证明的问题，并邀请数千名互不相识的读者尝试辨别哪些证明出自他们的程序，哪些出自真人（参见 https://gowers.wordpress.com/2013/04/14）。结果令人鼓舞，不过这当然不是图灵测试，因为选定问题的并非独立裁判。[26]1993 年，奔腾推出了一款芯片，后来发现其浮点处理器存在硬件缺陷（即使是在对某些七位整数进行除法时也会出现）。更准确地说

商业软件平均每百行代码就有一个漏洞；即使是航天计划，也只能将这个数字降到每一万行代码大约一个漏洞。企业为了限制法律责任，会对漏洞保持沉默，而修复漏洞还可能引入新的漏洞！近期出现了一些程序验证器，旨在检查程序是否执行了其声称的计算，因此错误或许很快会减少，但这还有待观察。

5. All the proofs yet to come…

本期其他文章探讨了计算机将预测出哪些定理，甚至探讨了计算机是否会成为普遍的思想领袖，但这类推测超出了我的范围。眼下，计算机证明验证与生成是发展迅速的领域，能见证这些重要进展的开端，令人振奋。许多杰出人士正投身其中，Kevin Buzzard 持有一个颇有说服力的观点：“越多人熟悉这款软件，有趣的事情就会越早发生”——如今，Lean 已有数千名用户。

我关注的是证明的本质及其未来走向，主要着眼于反驳一种天真的看法：形式化证明会提升客观性。但更重要的是，我们不应失去社群式证明方法的益处；这种方法长期以来让我们受益良多。事实上，我们的一些传统应当指引我们探索如何发展计算机证明。

最重要的是，计算机应当服务于我们的需求。输入应采用标准词汇表，证明应以人类可读的形式呈现，而验证者则应被开发出来，与数学研究者协同工作，让她能够更轻松地检验一个提出的证明思路。当程序声称已找到一个证明时，应由不同的计算机、使用不同的芯片和不同的软件进行验证。[27]

形式化证明非常脆弱（也就是说，如果我们发现任何错误，就会对更多内容产生怀疑），因此我们需要融入传统直观数学证明的稳健性，以便修正细小的缺陷，

最终，我们希望得到信息丰富的证明，而不仅仅是验证；这些证明应当解释数学，而不一定要给出最短的证明。这应当使我们能够修改所学内容，并帮助我们发展新的概念。我们应该能够像与一个（记忆力极佳的）人类互动那样，与计算机程序进行交互。对于同一个定理，给出多个证明可能是有益的，例如突出不同的主题——计算机程序能否经过训练来识别这些变体？[28] 很难描述什么是好的证明（因此也很难描述厄德¨什的“《书》”），但我们一看到它就知道：我们的计算机程序能否始终找到这样的证明？

我提出的要求很多，但现在无疑正是思考如何最好地引导这些发展、推动数学研究中证明的未来发展的时机。

他们没有召回那些有缺陷的芯片，而是保持沉默，并通过更新来纠正问题。然而，Thomas Nicely 在1994年6月发现自己在一次数论计算中算错了结果后，于是将这一错误公之于众。奔腾一直拒绝召回，直到 IBM 拒绝发运其产品。[27]这样一来，在每个系统中“证明”的同一点出现错误的可能性就会非常低，尽管这仍不能保证正确性。[28]我们可以问：给定一个直观证明，对它进行形式化后得到的还是同一个证明吗？当一个直观证明被拆解为 Lean 运作所需的各个部分时，它看起来会非常不同，并且依托于一个外观相当不同的知识库。同一个证明在针对另一种语言进行修改后，又会显得有多么不同？

References

[1] Kevin Buzzard, *The Xena Project*, 博客：https://xenaproject.wordpress.com/ [2] 尤金妮娅·郑, *Mathematics, morally*, 2004年（预印本）。[3] 亚历克斯·戴维斯、Peter Veli˘ckovi´c、乔迪·威廉姆森等, *Advancing mathematics by guiding human intuition with AI* 600 2021 70 74 , 《自然》（），–。[4] 西尔维娅·德·托福利, *Reconciling Rigor and Intuition*, 《认识》（2020年）。[5] 西尔维娅·德·托福利, *Groundwork for a fallibilist account of mathematics*, 《哲学季刊》71（2021年），823–844。[6] M. 加内萨林甘和W. T. Gowers, *A fully automatic problem solver with human-style output* 58 2017 253 291 , 《自动推理杂志》（），–。[7] Kurt G¨odel, *The modern development of the foundations of mathematics in the light ofphilosophy*, 1961年向美国哲学学会所作演讲的草稿，牛津《文集》第三卷（1995年），375–388。[8] 安德鲁·格兰维尔, *Accepted proofs: Objective truth, or culturally robust?*, 《数学与哲学年刊》（2023年），但愿如此。[9] 唐娜·哈拉维, *Situated knowledges: The science question in feminism and the privilege of partial perspective* 14 1988 575 599 , 《女性主义研究》（），–。[10] 托马斯·S·库恩, *The Structure of Scientific Revolutions*（第二版），芝加哥, 芝加哥大学出版社，1962年。[11] 帕特里克·马索, *Formal mathematics for mathematicians and mathematics students* YouTube讲座（https://www.youtube.com/watch?app=desktop&v=tp h3vzkObo）。[12] 梅兰妮·米切尔, *Artificial Intelligence: A Guide for Thinking Humans*, 法勒、施特劳斯和吉鲁出版社，2019年。[13]罗德里戈·奥奇加梅, *Automated mathematics and the reconfiguration of proof and labor*, 预印本。[14] 朱利安·赖斯和扬·施普伦格, *Scientific Objectivity*, 《斯坦福哲学百科全书》（2020年冬季版），爱德华·N·扎尔塔（编），https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=scientific-objectivity[15] Endre Szemer´edi, *On sets of integers containing no k elements in arithmetic progression*, 《算术学报》27（1975年），199–245。[16] A. M. 图灵, *Computing Machinery and Intelligence*, 《心灵》新系列59（1950年），433–460。[17] 阿克谢·文卡特什, *How we place value in mathematics*, 预印本。

Département de Mathématiques et Statistique, Université de Montréal, CP 6128 succ Centre-Ville, Montréal, QC H3C 3J7, Canada.*Email address*: andrew.granville@umontreal.ca