C

SCP/38/4
原文：英文
日期：2026 年 9 月 28 日

专利法常设委员会

第三十八届会议
2026 年 11 月 2 日至 6 日，日内瓦

有关专利和新兴技术的背景文件（SCP/30/5 更新）

秘书处编拟的文件

目 录

一、导 言 3
二、人工智能技术概况：神经网络、深度学习和生成式人工智能 4
A. 机器学习 4
B. 神经网络 5
C. 深度学习 7
D. 生成式人工智能 9
E. 目前深度神经网络、生成式人工智能和代理式人工智能的局限 19
F 人工智能与创新的当前技术格局 21
三、人工智能相关发明的专利保护 22
A. 总论 22
B. 可授权主题 25
C. 新颖性和创造性 33
D. 评估可专利性背景下的人工智能生成的现有技术和 PHOSITA 35
E. 充分公开和权利要求 40
F. 工业实用性 45
G. 发明人权利和所有权 45
四、作为工具应用于专利制度的申请过程和管理工作的人工智能技术 48
A. 知识产权主管部门的工具 48
B. 申请人、第三方和知识产权专业人员的工具 49

附件[第 1-3 页]

一、导　言

1.　在 2018 年 12 月 3 日至 6 日在日内瓦召开的专利法常设委员会（SCP）第二十九届会议上，委员会商议秘书处将编拟一份有关专利和新兴技术的背景文件并提交 SCP 第三十届会议。因此，秘书处向 SCP 第三十届会议提交了文件 SCP/30/5 以供讨论。

2.　在委员会第二十九届会议的讨论中，很多发言的代表团都提到人工智能是委员会待讨论的问题。因此，尽管“新兴技术”一词可能有着宽泛的涵义，涵盖了多项不同的新技术，包括人工智能和机器学习、区块链、合成生物学及基因编辑等，但文件 SCP/30/5 涉及的是有关专利和人工智能的背景信息。

3.　在 SCP 第三十五届会议上，委员会商议秘书处将更新文件 SCP/30/5，根据从成员国和地区专利局收到的信息，汇编与人工智能相关发明可专利性有关的现行法律和实践，并提交 SCP 第三十六届会议。根据该决定，文件 SCP/30/5 第三部分 A 节至 E 节得到更新并提交第三十六届会议（SCP/36/5）。

4.　此外，为保持文件的完整性和可读性，秘书处在文件 SCP/30/5 第二部分中增加了对生成式人工智能的概述。此外，根据与专利专题有关的 SCP 活动，更新了文件 SCP/30/5 第三部分 F 节和第四部分并提交 SCP 第三十六届会议供委员会审议。

5.　本文件保留了文件 SCP/30/5 的结构，加上对文件的相关更新（SCP/36/5），由三部分组成。文件第一部分提供了关于人工智能技术的背景信息。文件第一部分阐述了人工智能技术的基本概念，尤其是构成当前人工智能发展核心的机器学习技术和生成式人工智能。这种对于技术的概述性说明被认为是必要的，因为为了了解一项技术对于专利制度的影响，至少要对这项技术本身具备基本的了解。在此过程中，简要概述了自上次更新（SCP/36/5）以来人工智能领域的最新技术进步。这些进步涉及深度学习系统效率和部署方面的最新变化、多模态生成式人工智能，以及基础模型和后训练中的变化，以及代理型人工智能。当前对本文件第二部分的更新增加了新的 F 节，该节基于产权组织人工智能与创新工具包对人工智能和创新格局进行了梳理。本文件第二部分和第三部分阐述了专利制度和人工智能之间的相互关系。它们处理了两个不同的问题：第二部分探讨了作为专利保护客体的人工智能技术（或人工智能相关发明），第三部分讨论了专利制度的主管部门和用户对于人工智能技术作为工具加以应用。已记录产权组织成员国报告的以及基于秘书处开展的独立研究的变化。第三部分新增的 D 节讨论了在评估可专利性背景下，人工智能生成的现有技术和“本领域普通技术人员”（PHOSITA）方面新出现的问题，因为这些挑战正越来越多地在专利法和政策框架下得到讨论。相应地对第三部分进行了重新编号。

6.　关于“专利质量”一词，尽管无法为它找到唯一的定义，但通过 SCP 此前的活动得到了两个主要概念。它们是：(i)专利本身（即符合实质性标准并被授予专利的发明）的质量；及(ii)专利局和专利局以外的专利程序的质量（文件 SCP/27/4 Rev.）。从这个角度来看，可以说有关人工智能相关发明专利保护的问题涉及专利质量的第一个方面，而通过 AI 技术改进专利程序方面的问题与专利质量的第二个方面有关。

7.　此外，本文件载有一份列有世界知识产权组织（产权组织）和各知识产权局有关人工智能和专利的网页的附件。为了便于本更新文件的读者查阅，新增或经修订的段落（或段落内的新增内容）和脚注均以蓝色标示。为支持视力障碍者的无障碍获取，这些段落为第 3 至 5 段、第 7、8、9、15、17、23、24、26、34、36 至 39、54、59、65 至 77、79、82、84 至 86、89、90、93-96、101、103、104、108、113、115 至 118、120 至 124、126 至 128、130、131、133 至 145、

150、153、156 至 173、182 至 184、188 至 200、209 至 211、213、214、216 至 219、223 至 225、227 至 231，以及附件第 1 至 3 页。

二、人工智能技术概况：神经网络、深度学习和生成式人工智能

8. 尽管对于人工智能（AI）尚无单一、普遍接受的定义，但定义了人工智能系统的法律和政策文书通常采用宽泛、技术中立的表述，这反映了人工智能技术不断演变和多样化的性质。在此背景下，欧盟《人工智能法案》和经合组织的定义都将人工智能系统描述为一种基于机器的系统，具有不同程度的自主性和适应性，从输入推断出如何生成输出，例如预测、内容、建议或影响实体或虚拟环境的决定。因此，出于技术介绍之目的，[1]人工智能系统主要可视为从数据中学习或推断模式的系统。

9. 目前，仅少数司法管辖区提出了对人工智能及相关概念的定义。举例而言，大韩民国在其《2025 年关于人工智能发展与创建信任基础的框架法》中有“人工智能”、“人工智能系统”和“人工智能技术”的定义。“人工智能系统”一词被定义为“基于人工智能的系统，以不同程度的自主性和适应性，出于给定目的推断出影响现实和虚拟环境的输出，如预测、建议和决定”，而人工智能本身则被定义为“以电子方式实现人类学习、推理、感知、判断和语言理解等智力能力”。[2]尽管这些定义主要在监管领域而非专利法背景下制定，但体现了对人工智能技术最新发展的全球广泛共识和理解。不过，需指出，尽管监管定义可以作为理解人工智能技术本质的良好起点，但就专利法目的而言，围绕人工智能发明的具体术语需要更深入且细致入微的分类（见下文第二部分第 2(F)节）。

10. 本文件第一部分介绍了神经网络（NN）、深度学习（DL）和生成式人工智能领域最重要的技术概念，它们是当前最为重量级的人工智能技术。[3]本节使非计算机专业人士得以了解这些新兴技术如何便利地发挥作用，以帮助人们更好地了解 AI 技术和专利之间的相互关系。

A. 机器学习

11. 从历史上来看，首批 AI 方法是对机器编程。编程在这里是指人向机器发出步进指令以完成一项任务。例如上世纪 80 年代的主要 AI 方法是专家系统，它通过各领域专家所撰写的规则复制人类的专门知识。由于这些方法成本高且具有局限性，随之而来的是发生在 1987 年至 1993 年期间的所谓第二次 AI 冬天。

12. 相比之下，机器学习（ML）方法探索的是机器如何通过输入示例和预期输出学习一项任务的解决方法，同时无需以步进指令的顺序对如何实现目标直接进行编程。这一方法更接近实际的生物性认知：儿童学习辨认物体（如杯子）的方法是通过相同物体（如各种不同种类的杯子）的示范作用。它是到目前为止最为主要和成功的 AI 方法。

---

1 见制定关于人工智能统一规定的欧洲议会和理事会 2024 年 6 月 13 日第 2024/1689 号条例（欧盟）（《人工智能法案》）第 3 条第(1)款；经合组织，“关于经合组织人工智能系统更新定义的说明备忘录”，《经合组织人工智能论文集》第 8 期，经合组织出版，2024 年，第 4-5 页。这两种定义均使用基于机器、技术中立的术语，并提及自主性/适应性以及预测、内容、建议或决策等输出。

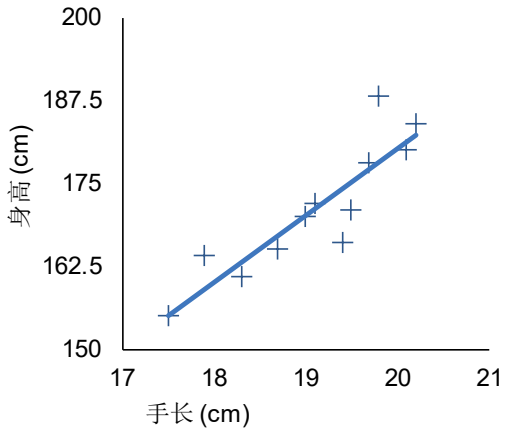
2 见《关于人工智能发展与创建信任基础的框架法》（2025 年 1 月 21 日第 20676 号法案），可访问：https://elaw.klri.re.kr/eng_service/lawView.do?hseq=71019&lang=ENG

3 2019 年世界知识产权组织技术趋势——人工智能，第 31 页。机器学习在 AI 相关专利申请中的占比为 89%，在所有 AI 相关专利中的占比为 40%。在 2013 年至 2016 年期间，深度学习在机器学习技术领域中的年增长率达 175%，神经网络增长率为 46%。

13. 一般来说，机器学习法接受观察结果的输入，并把它们用于预测输出。这种学习法会试图从给出的输入输出对数据集中建立起数学模型，使它的预测和预期输出之间的差别最小化。通过这种方式，它尝试学习给定输入和输出之间的关联/模式，这些关联/模式可被代入之前未见过的新输入。

14. 为了说明这一学习过程，我们可以看一看最简单的机器学习法——线性回归。假设我们希望学习如何把人的身高和他/她的手长相互关联。已有一些身高和手长对的观察结果（左表），在下图中用十字表示：

身高 (cm)	手长 (cm)
170	19.0
155	17.5
184	20.2
188	19.8
178	19.7
172	19.1
165	18.7
180	20.1
161	18.3
171	19.5
164	17.9
166	19.4



手长 =
0.098 * 身高 + 2.23

if 身高 = 180 cm,
手长 =
0.098 * 180 + 2.23
手长 = 19.9 cm

15. 线性回归是一种在这些点之间找到误差最小的直线的方法。使误差最小化的过程就是训练。数学方法通过找到距离数据点最近的直线来实现这一训练。一旦找到了这条误差最小的线，即可根据一个人的身高预测他/她的手长。例如，如果一个人身高 180 厘米，模型会预测其手长为 19.9 厘米（见右框）。在这个简化的示例中，方程中两个学习得到的数值，即斜率和截距，是该模型的参数。更贴近实际的机器学习系统也会从数据中学习参数，只不过数量多很多，数学函数也更复杂。

16. 这种简单的方法对于学习更复杂的问题来说无疑具有很大的局限性，例如涉及两个数值变量的问题。在上述例子中，看起来除了身高，还应考虑年龄和性别，才能使手长的预测结果更可靠。还会使用更复杂的数学模型，尤其是不仅限于直线的非线性模型。

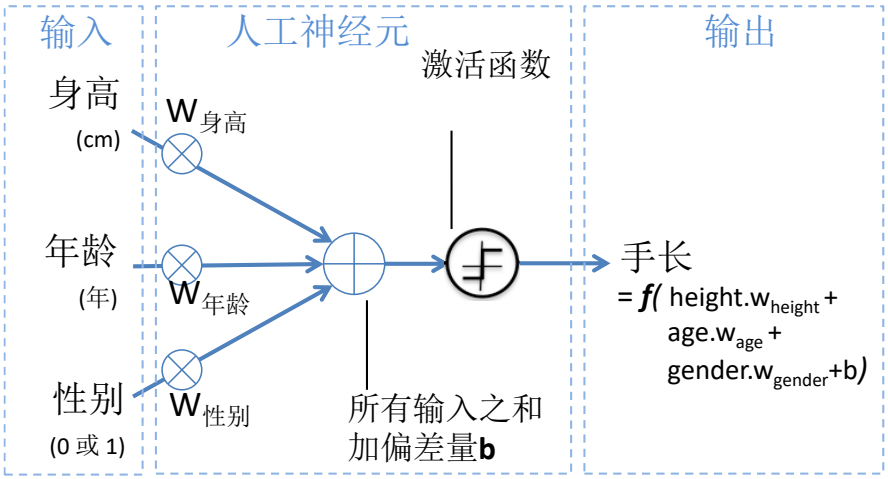
17. 在那些更为高级的方法中，神经网络（NN）提供一种通用预测器，能够接受任何类型的输入。神经网络尤其擅长解决涉及非结构化数据作为输入（如图像或语音、文本、代码、视频、音频，以及经过数值编码后的生物或化学数据）的任务。作为一种高级的神经网络类型，深度学习仍然是人工智能专利申请以及当前生成式人工智能系统中的核心技术。产权组织的 2024 年生成式人工智能专利态势报告指出，深度学习和计算能力的进步推动了生成式人工智能的发展，自 2017 年 transformer 架构问世以来，生成式人工智能专利族增加了超过 800%。

B. 神经网络

18. 支撑神经网络的基石是人工神经元，也被称为感知器或节点。它由弗兰克·罗森布拉特在 20 世纪 50 年代和 60 年代开发。神经元接受 n 个被称为特征的输入，它们是待处理数据（像素、

词语、信号等）的数值表示。每个输入乘以权重并求和（见下图）。经过加权求和的组合值与偏置量 b 相加。最后得到的数值传送至激活函数 f。

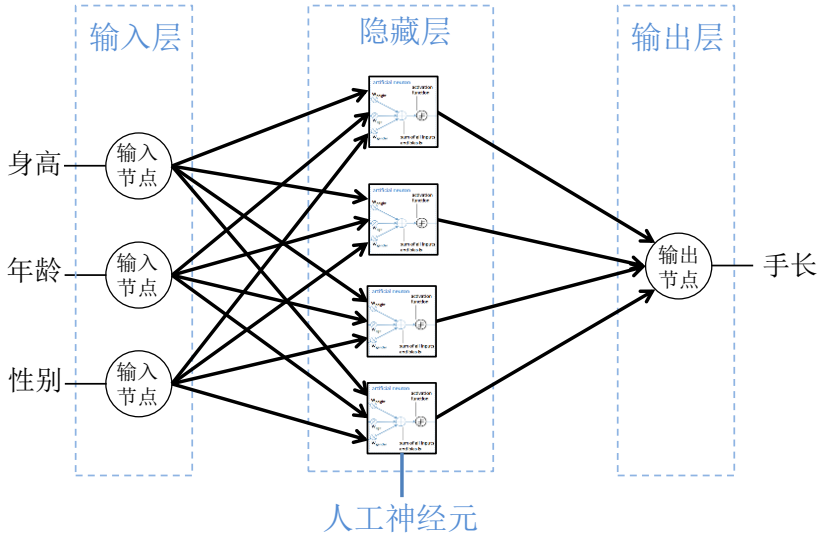
19. 例如，回到预测手长的例子中，如果已有一个人身高、年龄和性别的数据，则人工神经元表示如下：



20. 权重反映了相关输入特征的强度，也就是说一项特征就其本身而言对于最终结果产生多大的影响。

21. 激活函数模拟生物神经元的“放电频率”——传播最终信号或是不传播信号。它对经加权求和的输入进行某种固定的简单数学运算。最常用的激活函数是 ReLU（修正线性单元）。[4]

22. 人工神经元是一个相对简单的函数。对它进行编程可以使用少于 25 行的代码。那么一个完整的神经网络至少由三层组成：一个输入层、一个或多个隐藏层和一个输出层。输入层和输出层包含不进行计算的节点。它们仅仅是为输入节点向隐藏层传递数值信息，或是为输出层把信息从网络传送到外部世界。隐藏层包含如上所述的人工神经元。来自相邻各层的节点彼此之间存在用箭头显示的连接（或边）。



[4] ReLU 把一个数值作为输入，得到的是 0 与该数值中最大的那个数。例如，如果输入是“1”，则输出为“1”，如果输入是“-1”，则输出为“0”。

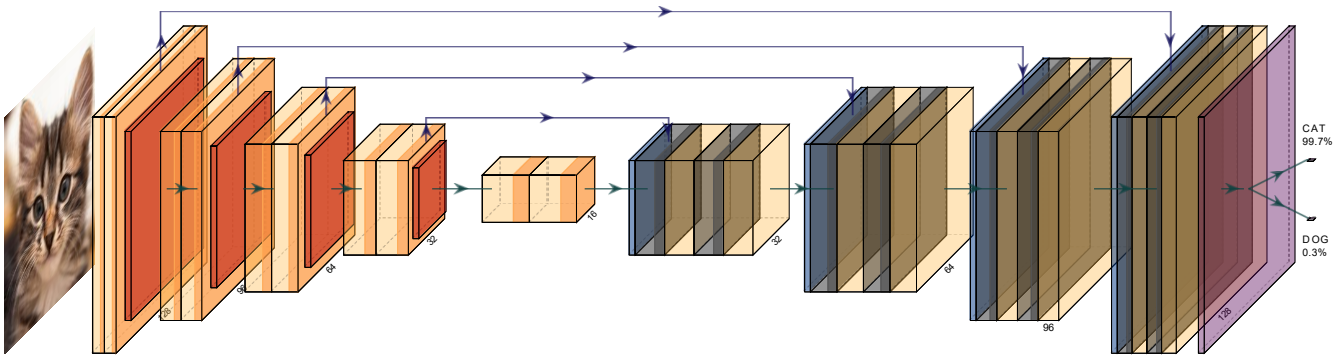
23. 输入层充满了数字编码信息，然后通过隐藏层向前传播。初始数值被隐藏层的神经元修改，然后传播至输出层并与最终输出相对应。输出节点的数量与预期从神经网络得到的应答数量相匹配。例如在这个例子中，预期得到的是手长这个唯一值。数据在这里总是经过各层向前移动。对于常见的前馈网络（例如用于分类或预测的多层感知机），推断是贯穿各层的单向计算。其他架构则以不同的方式处理序列信息。

24. 训练神经网络就是对隐藏层中所有神经元的权重和偏置参数进行设置，使在一组示例中所观察到的偏差最小化，这类似于上文 A 部分中所述的线性回归。训练神经网络的机制基本上是“从错误中学习”。训练数据由一些输入/输出对组成。当神经网络有了一个输入后，它对相应的输出可能是什么进行随机“猜测”。然后它会查看其应答与实际输出之间存在多大差距，并且再对其权重和偏置量作出适当调整。所有输入/输出对都要重复这个过程，直至得到最优权重和偏置量。在当前术语中，这种调整通常通过名为反向传播的算法计算，并结合梯度下降或其中一种变体等优化方法。

25. 应注意的是人工神经元只是在很小程度上受到哺乳动物生物神经元结构的启发，并且复杂程度比后者低得多。生物神经元相比人工神经元要复杂多样得多。有很多因素（突触结构和几何结构、神经递质类型等）会影响信号传播。例如突触包含超过 2000 种不同的蛋白质，具有非常多种多样的物化特性。[5]

C. 深度学习

26. 尽管神经网络从上世纪 50 年代开始就已为人知晓，但直到 2000 年代，隐藏层数量仍然通常仅为一个。自 2010 年代以来，计算能力的提升、更大的数据集以及增强的训练方法成为可能，使神经网络层数在过去 10 年中得以增加（由此“深化”）。例如对于猫或狗图像分类问题（图像中是猫还是狗？），目前的深度神经网络如下所示：



27. 在以上的例子中，我们看到与前述的简单神经网络相比，目前的网络在复杂程度方面发生了很大变化：

(i) 输入节点数量很多；每个输入节点接收一个图像的一个像素的信息。在以下示例中，猫的图像被转换成一个尺寸为 18*40 的矩阵，每个像素由其灰度级定义，灰度级由 0（白色）到 256（黑色）之间的整数表示。对于写实猫狗图像的分类，我们通常

[5] “人工和生物神经网络之间的区别”，Nagyfi Richárd，登于 Toward Data Science 上的博客文章，2018 年 9 月。https://towardsdatascience.com/the-differences-between-artificial-and-biological-neural-networks-a8b46db828b7。

使用尺寸为 128×128 像素的图像，每个像素由红绿蓝 3 个值确定，即 49,152 个输入节点，因此以下每个神经元有 49,152 个输入特征。

人对图像的感知



计算机对图像的感知

```
  5   4   6   5   5   5   6   6   7   9   8   9   8   4   6  11  40 136
  6   8   5   5   5   5   5   5   6   7   8   8   9   4  65  27  18  61
  5   9   7   4   4   4   4   5   5   6   6   7   7  70 123  86   9  26
  4  82 107   6   5   4   4   4   4   5   6   7   5  85  90 111   6  10
  4 103 132  77   4   3   4   4   5   5   6  10  26  91  98 127   5   7
  8 140 139  93  18  11   5   3   5   5   6   6  66  92 133 149   4   9
 10 149 102  78  73   6  10   4   5   5  11   8  91 120 122 160   9  16
 15 155 149  85  94   5  10   5   6   8  17  30  86 115 143 121   9  28
 22 147 149 142  80 101  27  18  23   7 140  85 119 137 157 129  13  46
 15 144 146 126 113 126 135  68 156  97 182 129 121 158  95 152  18  61
 10 150 149  99 135 144 217 212 176 119 190 174 112  83  94 142  36  81
 16 144 147  85 101 135 179 223 196 107 201 151 131  87 161 123  41 105
 43 110 114 115  68  93 147 187 218 197 229 165 102 136  99  98  74 117
 27 114 133  83  75 117 179 168 168 184 234 161 181 142 149  86  95 112
 11  84 105 121  84 153 188 183 215 171 228 198 169 161 124  56  84 109
  3  66 110 112 116 135 136 233 235 218 127 196 174  97 110 164 144 110
  2  48  84 105 113 148 149 238 183 146  74 195 187 150 173 204 160 102
  6  21 136 129 114 153  98 192 144 172  61 124 120 138 189 188 176 101
  9  11 180 140 121 122  92 244 133 207 107  70 102 117 192 147 172 112
  6   9 179 168 121  53  40 141 116 154 119 141 230 135 216 189 180 138
  8   9 165 185 100 125 242 104 133 117 138 193 204 126 223 191 193 175
  8  33 183 176 139 212 171 231 123 132  63 208 192  65 195 195 205 203
  7  12 181 233 209 127 210 188  51 164  40 197 192 184 200 217 211 216
  6  14 191 229 150  42 195 206  82 132 189  74 107 212 161 223 211 216
  8  22 202 191 185 139 181  65 170 105 136  78 187 199 174 170 207 218
 10  29 160 163 178 193 211  62 211  75 188 110 171 200 217 173 213 195
 10  21 104 137 158 158 158 113 193  95 156 161 134 208 184 171 206 201
 21  35  98 195 234 163  86 111 131 112 156  72  68 139 172 168 206 201
 31  46 101 161 159 155 186 151  53 107 195  93  67 114 190 167 197 204
 32  29  87 188 124  87  30 192  49 225 112  57  60 133 148 154 195 142
 40  26  66 104 194  67  36  97  48 152 127 106 119 203 178  72 185 181
 55  39  72 145 166 190 103 114  91 174 147 114 176 125 179 109 194 220
 97  72  89 173 163 185 163  91 102 114  94  98 164 156 163 114 184 224
135 111 111 187 160 143 169  94  63  76  84 115 144 155 169  89 128 215
163 151 115 164 184 174 126 135  74  86  90 132 150 143 123 150 187 216
195 173  88 131 193 183 112 109 106 104 120 128 114 131 158 175 191 213
199 177 178 185 158 127 112 114 101 104 112 123 167 166 163 197 187 207
139 157 161 116 126 122 124  89  92  89 110  97 142 169 189 210 174 194
121 112 140 122 141 110  92  90  85  80  94  91 156 214 211 152 139 185
 97 104 101 170 153 114 106  95  88  96  84 129 217 221 157 156 178 205
```

(ii) 多层神经元被用于连续处理输入信息。对于图像处理，10 层以上的神经元并不鲜见，每层可能包含上百个神经元，它们通常有不同的排列方式以提供特定的优势。

(iii) 这种典型的深度神经网络可能有数千万个要在训练期间设置的权重和偏置量，需要上万个有标号图像。

28. 令人惊讶的是在像 Keras[6]这样的开源（OS）框架中，经过培训的数据科学家可以用少于 100 行的代码实施深度神经网络。在处理一个猫狗图像在线开放数据集的情况下，使用商用硬件的网络在分类方面的准确性可达 93%：这与人的表现相差无几（对于这类任务人的准确性估计约为 95%）。

29. 层数的递增使全球预测任务所涉及表示和过程的层级这一概念应运而生。第一批层通常捕捉输入数据中的低层级模式（就处理图像而言如线条、彩色区域等），中间层识别的是更高层级的结构（如猫狗分类情况下典型的猫耳或口鼻），最后的层自我分工，依据所识别的结构完成最终的预测任务。

30. 与传统的神经网络相比，深度神经网络具有若干重要特征，这也是后者在当前取得成功的原因。

发现特征表示

31. 传统的机器学习使用工程师人工编制的特征来解决问题。例如，对于预测手长，机器学习工程师要根据自己的直觉和尝试选择一些特征，如人的身高、性别和年龄。这个步骤被称为特征工程。一个特征是机器学习算法用于预测输出的数据的一个方面。这一步普遍十分耗时，如果处理的是非结构化数据（图像、文字、语音、视频），效率还相对较低。

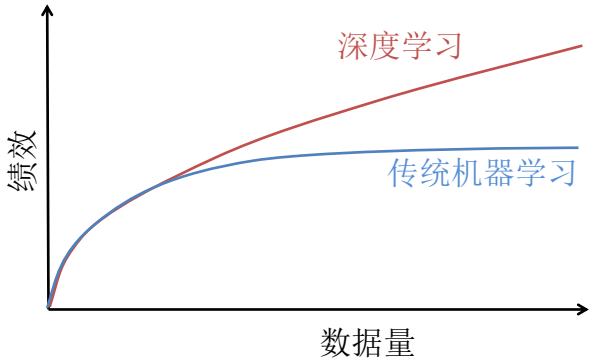
32. 纵观机器学习的发展史，深度神经网络首次显示出具有能够从原始数据中自动发现此类特征的实践能力。通过深化层数，神经网络在学习有用特征的同时还学习如何通过它们处理任务。例如在预测手长的例子中，操作人员只是向深度神经网络输入尽可能多的生物识别测量结果，让

[6] Keras：基于 Python 的深度学习库，François Chollet 等，2015 年-2024 年。https://keras.io/。

网络自动识别可用来作出最终遴选的数据。在图像分类的例子中也是如此，原始像素数据被发送给网络，后者对可作为判定依据的模式（如耳朵的形状、舌头或牙齿）进行识别，以判断输入的图片是狗还是猫。

数据规模和深度学习的绩效

33. 传统机器学习技术的绩效随着训练数据量的增加很快进入稳定期。这意味着增加更多的训练数据是无用功，训练算法经过一段时间后变得有些“饱和”。深度学习的重要特征之一是绩效随着训练数据的增加持续提升。这一特征解释了为什么在机器视觉中目前已有的最大网络可使用多达 1500 万个图像用于训练。



34. 从数学的角度来看，可把人工神经网络模型理解为仅仅是一组矩阵运算并找到导数。[7]随着计算能力的加强，深度学习可超过任何其他机器学习法，前提是有可供使用的海量训练数据和计算能力。

35. 我们能在多大程度上扩展这种神经网络，并继续提高深度学习模型的绩效和能力？为探索和提高这种扩展能力而开展的实验使生成式人工智能领域在当前取得了巨大成功。

深度学习模型的效率和部署

36. 近年来，第二条相辅相成的发展路径侧重于使深度学习模型更小、成本更低且更易于部署。剪枝、量化、蒸馏、稀疏激活模型以及参数高效微调等技术，能够减少模型所需的计算量或内存。例如，剪枝移除了神经网络中重要性较低的权重、神经元或其他结构，使所得模型在推断时所需的内存或计算量更少。[8]

37. 因此，深度学习系统并非总是作为在中央计算中心运行的大型模型部署。根据任务的不同，系统可能使用大型云模型、较小的专用模型、设备端模型，或模型和工具的组合。这些部署选择会影响用例、成本、速度、准确性、保密性、能耗以及人工智能系统的实际可获取性。

D. 生成式人工智能

38. 生成式人工智能可以定义为一种能够在用户的提示下创建新的、有意义的、类似人类的内容，包括文本、图像、音频或视频。“提示”此处指的是通常由人类用户生成的自然语言文本指令，尽管在最新的系统中，提示也可包括图像、音频、视频、代码、检索到的文件，或其他软件

[7] 对于向量处理器，可对这种数学计算进行高度优化（针对大量数据点多次重复进行相同计算），以及通过 GPU（图形处理器，与用于加速电子游戏的装置相同）或新的专门硬件把计算速度提高几个数量级。

[8] 见“深度压缩：通过剪枝、训练量化与霍夫曼编码压缩深度神经网络”（2015 年）；Tailin Liang 等，“用于深度神经网络加速的剪枝与量化：综述”（2021 年）。

工具或人工智能组件生成的输出。通过将文本与图像、音频、视频或其他输入相结合，提示可以为模型提供更广泛的背景，更接近人类的感知和数据处理，因为人类正是通过整合来自多个来源的信号理解世界。[9]

39. 下一节将介绍生成式人工智能的主要概念：

- 判别任务和模型与生成任务和模型之间的区别；
- transformer 架构和注意力机制，它们是现代生成式人工智能兴起的基础；及
- 大型语言模型（LLM）和基础模型的主要特点；
- 基础模型、后训练以及指令微调和偏好微调的作用；
- 检索、工具使用和代理式工作流程；和
- 能够处理和生成不同类型数据的多模态生成式人工智能系统。

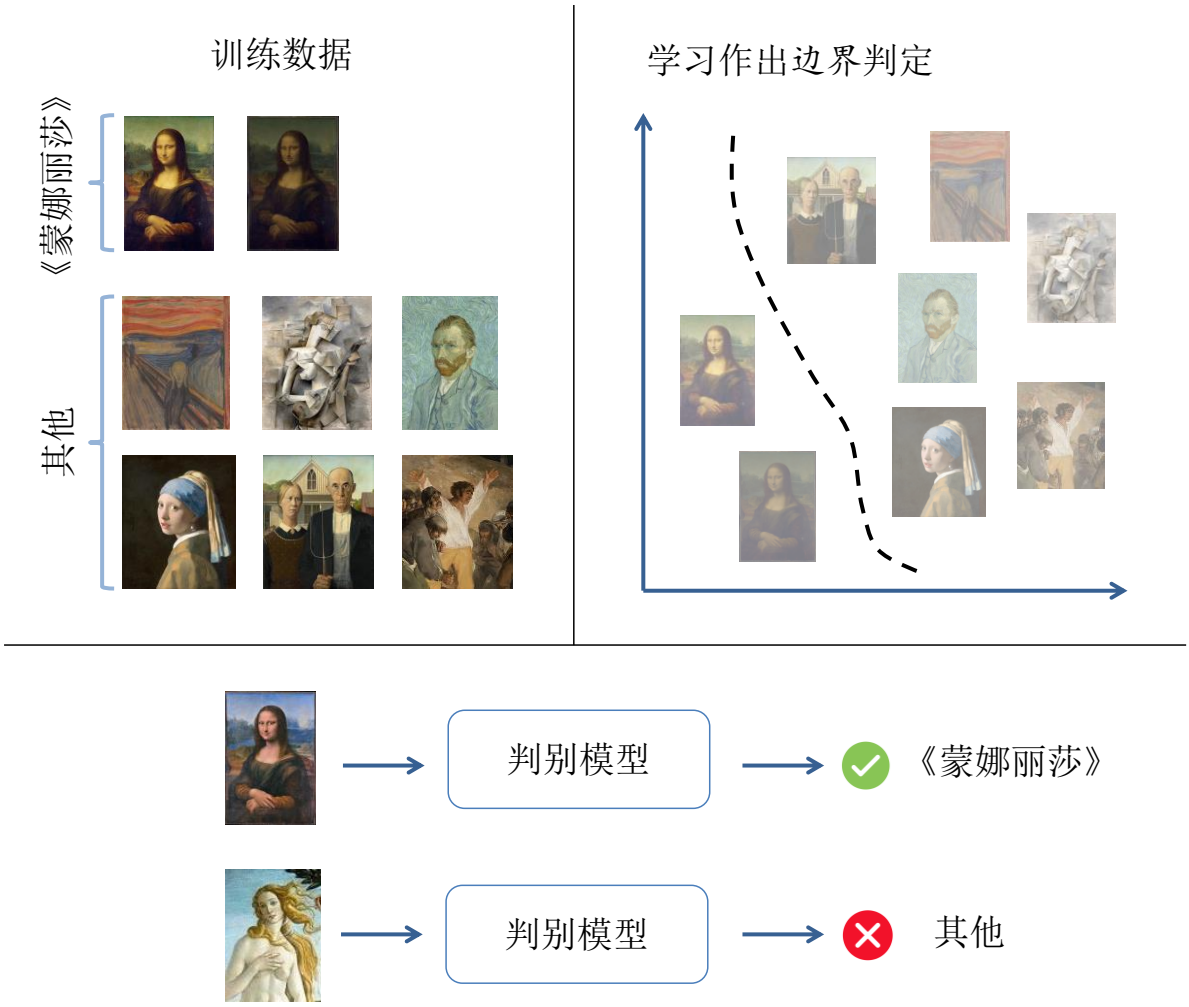
判别任务与生成任务

40. 深度神经网络通常可适用于两种不同的任务：

- 判别任务是对输入数据进行判定的任务，如分类、识别文本中的名称或分割图像。判别模型是为将输入数据分为这些不同类别而调整和训练的模型。
- 生成任务是指在给定输入数据的情况下创建新数据的任务。生成模型是为创建此类新数据而调整和训练的模型。它们通常用于翻译文本、生成图像、总结文本或回答问题。

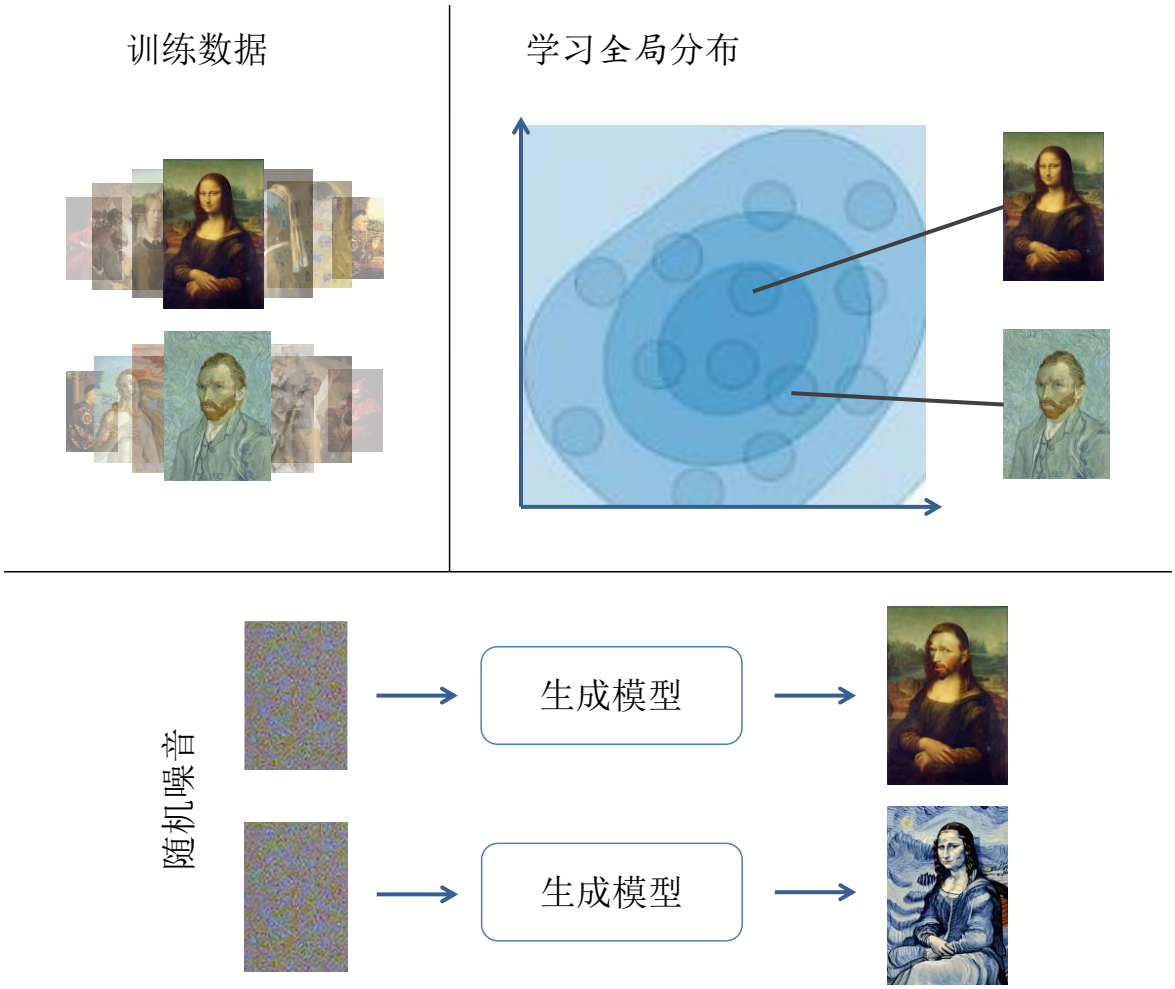
41. 以下两张图说明了机器学习任务的这两种基本类型。

[9] 见，例如 Tadas Baltrusaitis、Chaitanya Ahuja 和 Louis-Philippe Morency，“多模态机器学习：综述和分类”（2017 年），arXiv:1705.09406；Yuan Yuan、Zhaojian Li 和 Bin Zhao，“多模态学习综述：方法、应用与未来”，《ACM 计算概观》，第 57 卷第 7 期，文章编号 167（2025 年）。



来源：绘画图形来自公有领域下的维基共享资源（Wikimedia Commons），原图来自 2024 年产权组织《生成式人工智能》专利态势报告。https://doi.org/10.34667/tind.49740

42. 上图是对一幅图像是否为绘画作品《蒙娜丽莎》进行分类的判别任务。该模型必须学会如何区分两类图像：《蒙娜丽莎》和其他绘画作品。对于这类模型，学习过程主要集中在区分类别的标准上。因此，如果以绘画特征的空间来表示模型，该模型重点在于表示两类绘画之间的边界。



来源：绘画图形来自公有领域下的维基共享资源（Wikimedia Commons），原图来自 2024 年产权组织《生成式人工智能》专利态势报告。https://doi.org/10.34667/tind.49740

43. 上图是**生成新绘画样本的生成任务**。在生成任务中，模型必须学习每幅绘画的全局特征，以便生成连贯的新绘画。对于这类模型，学习的重点是表示绘画特征的全局分布。生成这两幅图像使用的是原版 Stable Diffusion[10]模型。

44. 判别模型擅长分类，但不能生成新数据。相比之下，生成模型也能处理判别任务，但准确率通常低于判别模型。与判别模型相比，生成模型的参数更多，计算成本更高，需要的训练数据也更多。因此，判别任务最初更受关注。随着算力的提高和可用的训练数据越来越多，生成任务变得更加可行，并在过去几年吸引了更多的开发活动。

Transformer 架构

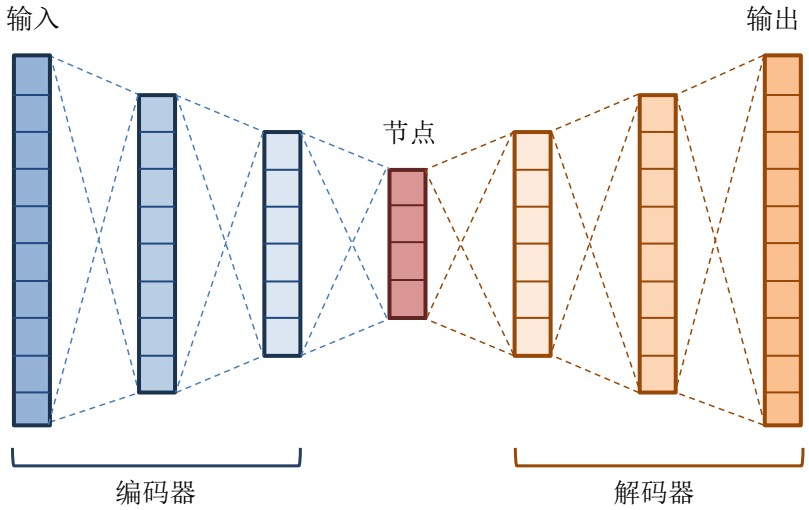
45. Transformer 是一种专为自然语言处理（NLP）任务设计的深度神经网络。它于 2017 年首次被提出[11]，与其他传统深度学习方法相比，它经过开发在训练过程中具有更好的扩展性。Transformer 基于两个重要概念：编码器-解码器和自注意力。

[10] Rombach R.、Andreas Blattmann A.、Dominik Lorenz D.、Patrick Esser P.、Björn Ommer B. (2021)。《使用潜在扩散模型进行高分辨率图像合成》。arXiv:2112.10752，https://arxiv.org/abs/2112.10752。

[11] Vaswani A.、Shazeer N.、Parmar N.、Uszkoreit J.、Jones L.、Gomez A.、Kaiser L.、Polosukhin I. (2017)。《注意力就是你所需要的一切》。神经信息处理系统进展 30。

46. **编码器-解码器**模型基于三个部分：编码器、代码和解码器。编码器是一个神经网络，它学习如何将输入数据编码并压缩成中间表示形式，也就是基本上为一串数字的代码。然后该代码在解码器中使用，解码器是另一个神经网络，它学习如何解压缩数据并将其重构为预期的输入格式。除了数据压缩，编码器-解码器的目标还在于学习如何表示某些数据的性质，以便对这种内部表示法进行修改后仍能重新构建出有意义的新输出。

47. 这种方法尤其针对机器翻译而开发，在机器翻译中，源语言的输入被编码为压缩的数字表示，然后根据大量的翻译输入/输出示例将其解码为目标语言。



48. 上图是编码器-解码器神经网络架构。输入（如文本或图像）通过若干层神经网络编码成向量。编码是一种中间压缩数字表示法，解码器可使用编码以所需的目标语言或图像表示法生成输出。需要注意的是，现代模型中的层数可能比上图显示的要多得多。

49. 对于专门用于判别任务的系统，通常会移除解码器部分，因为一般来说只输出分类，而不是新创建的数据。BERT（Transformer 双向编码器表示）模型系列就是这种情况，它是 2018 年至 2022 年期间机器学习领域的关注焦点。对于专门用于生成任务的系统，编码器可能会被移除，因为输出应该是新生成的数据，而编码器可能没有帮助。OpenAI GPT（生成式预训练 transformer）模型就是这种情况。

50. **注意力**是一种神经网络机制，能让学习过程同时关注文本的不同部分。它可以捕捉文本中的长距离依赖关系，这对于理解复杂的语言并对其进行表述非常重要。在实践中，注意力会根据重要性为输入词分配权重，从而使模型更加重视相关词。然而，如何选择这些重要的词呢？根据这一选择标准，存在着不同类型的注意力机制。

51. 最初的 transformer 架构采用的是自注意力机制：根据这种方法，会对训练数据所有输入中每一对可能出现的词之间的相关性进行评估。让我们快速了解一下这个过程：如果两个词在训练数据中出现的频率比随机出现的频率更高，那么它们的权重就会比不相关的词更高。

52. 总之，**Transformer** 是一种编码器-解码器深度神经网络，它利用注意力的优势，随着层数的增加，其训练时间比以前的深度神经网络更快。扩展能力与层数直接相关。堆叠更多的层意味着神经元全局数量成倍增加，从而可以存储更多的学习参数。在 Transformers 中，每一层都使用

注意力。因此，学习过程不会对输入进行多次循环，这种循环会造成很高的计算成本。相反，Transformer 会让输入经过多个注意力层。

53. 如下表所示，随着时间的推移，Transformer 的层数不断增加，从而可以在更大的数据集上进行训练，学习更多的参数信息。

模型名称	开发者	年份	层数	参数数量	训练数据量（估计数）
Transformer（原版）	谷歌	2017	6	1.1 亿	8 亿单词
BERT	谷歌	2018	12	1.1 亿	33 亿单词
GPT	OpenAI	2018	12	1.1 亿	6000 亿单词（40GB）
GPT-2	OpenAI	2019	48	15 亿	8 百万网页
GPT-3	OpenAI	2020	96	1750 亿美元	570GB 文本数据
GPT-4	OpenAI	2023	120	1.8 万亿（估计数）	13 万亿词

54. 该表格有助于说明 transformer 模型规模的历史增长。但不应将其视为当前领先模型的列表。对于许多最新模型，参数的数量、层数或训练词元并未向公众公开。此外，当前的能力还受其他因素影响，例如上下文窗口长度、支持的模态以及工具获取。

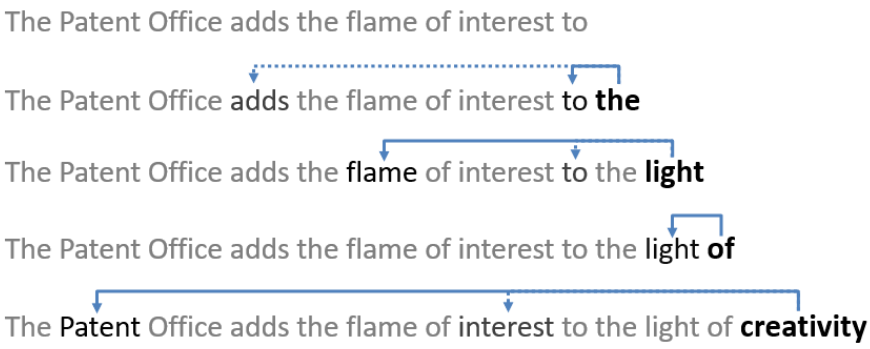
55. 我们之前提到，模型的生成能力会随着模型规模和训练的增加而提高。在这里，我们明白了为什么 transformer 架构使生成式人工智能得以兴起，而判别任务却位列最成功的人工智能系统之前。

从大型语言模型到 ChatGPT

56. 如上表所示，大型语言模型（LLM）是使用海量量训练数据训练出来的超大模型。它们的开发需要大型现代化计算基础设施。即使是拥有十万个处理器的大型计算中心，此类 LLM 的训练通常也需要两周到三个月的时间。这些模型是众所周知的 OpenAI ChatGPT、微软 Copilot 或谷歌 Gemini 等现代助手的基础。它们共享一种称为自回归的特殊生成机制。

57. **自回归模型**不是一次性生成完整的文本或图像，而是根据之前生成的要素，一次生成一个要素的数据。例如，对于文本，生成过程是一个迭代过程的结果，即根据之前生成的单词预测下个单词，直到预测出文本的结尾。这可以看作是模仿人类语言的自动完成功能。这也与众所周知的智能手机键盘的下个词预测功能类似。不过，这里所考虑的模型要比手机应用程序的下个词预测大得多，通常是后者的 1000 倍或更多。

58. 这种方法的一个主要优势是，对下个词或字的预测可以通过普通文本来学习。由于在训练过程中会扫描文本以设置模型的内部参数，因此这些模型有能力在给定文本上下文的情况下预测最有可能出现的下个词。这种训练不需要人工标注数据。计算机可以在没有人工指导的情况下学习识别模式，这种学习方式被称为无监督学习。然而，要达到这些模型目前的绩效，需要海量文本。



59. 上图是通过自回归生成的文本示例，说明了什么是注意力机制。当模型逐字生成文本时（粗体字），它可以通过注意力机制（蓝色箭头）关注与每个新生成的单词相关的先前单词。在实践中，可以达到的字词窗口规模非常大，在领先的模型中超过一百万。

60. 这种模型有两个有趣的特性：

- 它们不仅要学习一般语言，还要学习如何在训练数据中生成与世界上的实体和事件有关的各种事实的文本。
- 它们能够记住如此之多的文本内容，以至于如果提供足够的初始上下文，完成的内容就会变得出奇地精确和相关。

61. 这两个特性使得像 ChatGPT 这样的多功能助手成为可能，它们不仅能生成语法正确的文本。这些助手还可以回答问题，并实现各种任务，如总结、分类、改写、翻译、创作故事等。然而，考虑到这些模型当初的设计初衷只是为了预测下个单词以完成输入文本，它们如何能完成如此复杂的任务呢？

62. 这种协助用户和与用户对话的能力是通过用成千上万的**指令和预期结果**示例来调整这些模型来实现的。这些指令与用户为表达其要求而输入的“提示”相对应。模型将学习生成与预期结果相匹配的文本，以完成这些指令。基础 LLM 具备生成大量不同人类语言文本的能力，而这种额外的调整则训练模型学会如何使用这种能力来按照指令执行任务，从而实现流畅而多变的对话。

63. 以往传统的“监督式”机器学习模型需要大量特定任务注释的训练数据。相比之下，这些 LLM 只需写入自然语言“提示”即可生成新内容。因此，使用这些生成式人工智能工具不再需要技术技能。普通大众可以直接使用现代尖端人工智能，只需用自然语言指令提示模型即可。

64. 目前已开展了大量研究工作，以处理多个业务领域的各种任务和协助，以及生成其他语言形式，如编程代码或描述蛋白质的结构。有关应用和实例的综述，请参阅最近发布的产权组织生成式人工智能专利态势报告。[12]

[12] 世界知识产权组织（产权组织）（2024 年）。《生成式人工智能》。《专利态势报告》。日内瓦：产权组织。https://doi.org/10.34667/tind.49740。

基础模型与后训练

65. “基础模型”一词通常指在广泛数据上训练，通常使用大规模自监督且能够适应范围广泛的下游任务的模型。并非每个生成式人工智能系统都从基础模型构建，但许多当前系统使用一个经过广泛训练的基础模型，随后针对特定任务或领域进行调整。[13]

66. 需注意，权重和训练细节不对外公开的专有或闭源模型与模型权重向公众提供用于审查、微调或独立部署之目的的开放模型之间存在区别。对于非专业人士，当前许多生成式人工智能系统的开发可描述为三个阶段。第一，基础模型在广泛的数据上进行预训练。第二，对模型进行后训练，使其行为更实用、可控并符合部署目标，使用的技术包括监督式指令微调、基于人类反馈的强化学习（RLHF）、基于人工智能或自动化反馈的强化学习（RLAIF/RLAF）以及直接偏好优化（DPO）。通过这些技术，开发人员利用示范、排序、奖励或其他反馈信号引导模型行为。这意味着部署的产品并非仅仅是从预训练数据中产生中性输出的基础模型；后训练体现了有意的设计选择，通常用于提高准确性、安全性、透明度和问责制，并降低偏见或其他伦理风险。

67. 第三，部署的系统可与外部数据、工具或人工审查流程相连接。这些阶段并非总是严格分来，不同开发人员使用的术语各不相同，但这种区分有助于解释为何基于相似基础模型构建的两个系统可能在实际应用中表现不同。[14]

68. 近期的技术讨论还提及了迭代优化或模型辅助优化。在此类方式中，生成的示例、批评意见、评价或其他反馈信号可用于改进模型的后续版本，或优化工作流程中的输出。此类优化并不一定意味着模型会自主改变自身架构；在大多数实际场景中，开发人员或受控系统利用反馈、额外示例或评价结果更新模型行为或选择更好的输出。一项相关技术是知识蒸馏，即利用来自更大或能力更强的“教师”模型的输出或其他信号训练较小的“学生”模型。蒸馏有助于将教师模型行为的方面转移到更紧凑的模型中，从而降低部署成本、延迟、存储需求及其他资源需求，尽管学生模型也可能继承教师模型的局限性、错误或偏见。[15]

检索、工具使用和代理式工作流程

69. 一种无需重新训练模型而调整生成式人工智能系统的常见方法是检索增强生成（RAG）。在 RAG 中，系统从外部来源，如搜索索引、数据库或专利和科学文献集中检索信息，并将这些信息作为生成上下文的一部分提供给模型。检索到的信息可以使输出更针对具体领域且更易于检查，但结果的质量仍取决于检索质量、指令以及模型响应。[16]

---

13 R. Bommasani 等，“关于基础模型的机遇和风险”（斯坦福大学基础模型研究中心，2021 年）；NIST AI 600-1（2024 年 7 月）。

14 见 L. Ouyang 等，“基于人类反馈训练语言模型遵循指令”（NeurIPS，2022 年）（监督式微调和 RLHF）；R. Rafailov 等，“直接偏好优化：你的语言模型其实是一个奖励模型”（2023 年）（DPO）； Y. Bai 等，“宪法式人工智能：来自人工智能反馈的无害性”（2022 年）以及 H. Lee 等，“RLAIF：利用人工智能反馈扩展基于人类反馈的强化学习”（ICLR2024 年投稿）（RLAIF）。关于后训练方案及训练数据/方案的透明度，见 N. Lambert 等，“Tulu 3：推动开放式语言模型后训练的前沿发展”（2025 年）。

15 见 Chuanpeng Yang 等，“大型语言模型知识蒸馏综述：方法、评价与应用”，《ACM 智能系统和技术汇刊》，2025 年，DOI: 10.1145/3699518；Jianping Gou、Baosheng Yu、Stephen J. Maybank 和 Dacheng Tao，“知识蒸馏：综述”，《国际计算机视觉杂志》，129(6)，1789 - 1819，2021，DOI: 10.1007/s11263-021-01453-z。关于基础理论框架，见 Geoffrey Hinton、Oriol Vinyals 和 Jeff Dean，“神经网络中的知识蒸馏”，2015 年。

16 P. Lewis 等，“针对知识密集型自然语言处理任务的检索增强生成”，第 33 届神经信息处理系统大会（2020 年）。

70. 微调也是针对具体任务进一步训练模型的过程，其中权重基于在先知识分配。RAG 和微调是互补的调整策略，因为 RAG 更擅长处理需要对具体领域事实内容定期更新的任务，而微调则更适合调整格式或专门的推理模式。

71. 随着处理比简单文本和文件生成更复杂任务的需求出现，LLM 推理作为一种新型专用模型开发。这些模型能够建立逻辑联系以解决复杂的多步骤问题，而不仅仅是预测下一个单词。不同于直接生成答案，这些模型生成对应一系列工作流程步骤的中间提示。它们还能自我修正并评价自身的逻辑，在输出结果前重新审视中间提示。

72. 许多最新的系统还将模型与软件工具相连接，例如搜索引擎、计算器、代码执行环境、数据库、设计工具或应用程序编程接口。人工智能模型成为了“代理”，因为它们能够对外部服务开展行动，而非仅生成静态答案。“代理式人工智能”一词用于推理模型将目标分解为若干步骤，选择或调用工具，观察结果，并通过一次或多次迭代继续任务的系统。一个编排层或支撑性软件基础设施围绕模型和管理工具、存储中间结果，并在不同组件之间传递信息。在多代理工作流程中，多个基于模型的组件可协调或委派子任务。这些系统超越了独立的聊天界面，但其行为仍由模型、可用工具、指令、安全保障以及工作流程中内置的任何人工批准塑造。不同于每个阶段均需要人类提示的传统生成式人工智能，代理式人工智能能够基于待实现的目标和目的自主行动，规划、推理和记忆以实现复杂的多步骤目标，无需持续的人类监督。

专为生成发明定制的 LLM

73. 对于 SCP 来说，最相关的应用之一是**人工智能生成的发明**。为了实现这种能力，通常会对 LLM 进行微调，加入不同或额外的指令以及与发明相关的预期结果。此外，还可以利用专利文件、科学文本或其他技术内容对 LLM 进行进一步的预训练，并结合 RAG 来检索此类文献。

74. 这种定制模型可以有不同的用途：

(i) 模型用于帮助撰写专利申请，以一项人类发明的要点为基础，包括相关现有技术的建议。

(ii) 人类识别发明，并利用模型来识别这些发明可能解决的其他问题。

(iii) 人类识别问题，并利用模型生成可能的解决方案。

(iv) 人类利用模型识别潜在的有价值问题，并为这些问题生成可能的解决方案。

(v) 人类使用与工具连接的工作流程检索现有技术、生成候选解决方案、测试或模拟替代方案，并对结果排序，供进一步人类和基于人工智能的审查。

(vi) 人类在探索技术替代方案时，使用多模态系统将文本、图表、实验数据或设计文件结合起来。

75. 定制化 LLM 的不同用途可被视为逐步提高的发明人权利，从为人类发明人提供协助和见解，到更多地生成新构思。然而，还有其他因素决定了人类在生成创造性 LLM 输出方面所作工作的性质和程度：

(i) 创建用于查询模型的提示；

(ii) 为指导模型而提供的技术原则和特征；

(iii) 模型参数的选择；

(iv) 提示 LLM 的迭代次数。

(v) 系统检索到的外部文件以及与模型连接的工具，

(vi) 在选择、拒绝、测试或修改生成的替代方案时人类做出的决定，和

(vii) 提示的记录、检索的来源、模型版本、工具调用、中间输出及最终输出。

76. 从技术角度看，此类记录对于可复制性和可审计性至关重要，尽管此类日志的完整性和可靠性将取决于系统设计及其用户的做法。[17]

图像和其他模式的生成式人工智能

77. 我们在本文件中着重讨论文本生成。然而，在 LLM 之前就出现了成熟的图像生成模型，它们使用的是不同类型的模型。有意义的新图像的构建依据的是对十万张图像进行类似训练的模型，并由额外的机器学习机制进行控制。对训练图像的文本描述使文本和图像模式的结合成为可能。例如，扩散模型可以根据简短的文本描述生成高分辨率图像，2022 年发布的著名的 Stable Diffusion 模型就是一例。近期，多模态生成式人工智能已成为开发的标准方向，系统能够在单一工作流程中处理或生成文本、图像、音频、视频、代码或其他数据类型。



来源：Stable Diffusion 模型，提示为“一只阅读专利的猫”，原图来自 2024 年产权组织《生成式人工智能》专利态势报告。https://doi.org/10.34667/tind.49740

78. 上图中，图像生成使用的是原版 Stable Diffusion 模型，提示为“一只阅读专利的猫”。生成过程相当于：(i)提供带有随机噪音（随机像素）的图像，(ii)反复使用神经网络去除噪音。提供的文本提示用于驱动生成。随着噪音被逐步去除，利用通过大量训练数据学习到的与提示统计相关的图形元素就会构建出一个有意义的新图像。此类模型成功覆盖的其他模式包括视频（将一个图像连续转换为另一个图像）、语音、音乐以及由一幅或多幅二维图片构建的三维场景。

79. 在多模态系统中，提示不仅限于文本：图像、语音录音、视频片段、图表或其他输入也可能引导生成。在集成度最高的系统中（有时称为“任意对任意”系统），同一工作流程可接受一种以上类型的输入并产生多种类型的输出，例如，通过生成文本、代码、图表或进一步的语音响应回应关于一张图像的语音指令。

[17] NIST，《生成式人工智能和双用途基础模型的安全软件开发做法：SSDF 社区概况》，NIST SP 800-218A（2024 年 7 月）。

E. 目前深度神经网络、生成式人工智能和代理式人工智能的局限

深度神经网络是黑匣子

80. 与更为传统的算法相反，神经网络在学习过程中捕捉到的决策过程无法以人可以理解的方式直接表达出来。如上所述，深度神经网络可以自己学习数据中有用的特征。例如在猫狗分类的例子中，网络可以识别典型的猫耳或口鼻。但在实践中，这些特征大多数情况下无法被人解读。这些模式出自隐藏层中的数值优化过程，我们无法对其进行解读。

81. 此外，不可能以标准数学的形式显示对输入和输出之间关系进行定义的等式或系数。网络是最终的关系等式，可能涉及上千亿个参数。如此复杂的决策过程无法通过流程图或任何传统的算法表示法来说明。这就解释了为什么人们常常说深度神经网络是“终极”黑匣子，缺乏透明度。训练本身是由神经网络独自实现的，并且由此产生的网络极为复杂。

82. 当模型与检索数据库、外部工具或其他人工智能组件结合时，透明度可能变得更加复杂，因为观察到的结果不仅可能取决于内部参数，还取决于提示、检索到的材料、工具输出、模型版本以及部署设置。出于这一原因，系统工作流程的日志和文档记录已成为模型可解释性的重要实践补充。

深度学习需要很多数据

83. 一个令人惊讶的观察结果是，神经网络和深度学习就所涉及的数学建模而言属于最简单的机器学习模型。常见的说法是一名优秀的高中生就能理解其中的基本数学原理。但它们仍提供了到目前为止最好的结果。原因是它们对于利用很大规模的训练数据集最为适应。深度学习和生成式人工智能在当今的成功与其说是由于理论发展，不如说是由于计算能力的显著提升和可用的海量人类行为数据：我们通常称之为蛮力。

84. 深度学习明显存在的局限是蛮力不可能发生的情况。这其中尤其包括没有数据或数据有限的任务（如处理少见的人类语言、发现治疗罕见病的药物等），或是任务所在领域存在法律限制。现代生成式人工智能可通过重复利用基础模型，或以检索数据或合成数据补充稀缺数据集，减少一些障碍，但这些技术并不能消除对高质量、合法且具有情境代表性的数据的需求。

现实世界的数据存在偏差

85. 深度学习的成功取决于有大量的可用数据，但这种对于大量数据集的依赖也产生了若干问题：

– *数据偏差*：大规模的数据采集通常无法保证中立，一些群体在年龄、性别和民族方面的代表人数可能不足或过多。[18]偏差可能来自于数据采集法、已有的社会偏见或是创建数据集和模型的人员并非来自多元化群体。偏差还可能源于互联网规模数据中资源丰富的语言、地区或在线社区的过度代表，并且在汇总评价中可能不太明显。

– *偏差放大*：机器学习训练法天然地趋向于识别数据中的判别模式以快速提升预测或生成质量。因此，它们不仅学习我们实际存在的偏差，还经常将其放大。在生成式系统中，

[18] 亚马逊弃用对女性表现出偏见的秘密 AI 招聘工具，Jeffrey Dastin。路透社商业新闻，2018 年 10 月（ https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G）。

偏差放大可表现为刻板印象图像或文本、少数群体观点的排除，或输出的同质化，尤其是在将合成数据重复用于进一步训练时。

– 缺乏可再现性：由于模型依赖于专门的训练数据组成，再现一些所要求的结果只在非常少见的开放数据的情况下才有可能发生。

可信赖性

86. LLM 以令人信服的方式生成不准确内容的情况并不少见。这种不真实的输出被称为幻觉。尽管近一段时间以来专门开展了很多工作以降低此类事实错误的发生频率，但这类现象仍然很严重，尤其是在专业领域。类似的失误在多模态系统中可能出现，例如当系统错误解读图像、图表、声音或视频，并自信地表达错误时。问题不仅限于事实错误；还可能包括推理缺陷、引用不准确或缺乏依据的高度确信。

87. 最近的一项研究估计，在简单任务中，GPT-4.0 每次输出产生幻觉的频率约为 28.6%[19]，这可能高于用户通常的认知水平。例如，当将人工智能应用于法律领域时，斯坦福大学研究人员的一项研究[20]测定，GPT-4.0 每次响应产生幻觉的频率至少为 58%。根据同一组研究人员的一项研究，即使是专门使用法律文本进行预训练并使用法律数据库来丰富和控制提示（一种称为“检索增强生成”或“RAG”的技术）的定制化 LLM，如律商联讯和汤森路透的商业化 LLM，其每次生成响应的幻觉率估计仍在 17%到 33%之间。[21]

88. 对 LLM 进行长期评估似乎很困难。在经典的机器学习评估中，系统不应通过其训练数据进行测试和评估。这一原则确保了不同的系统可以使用开放访问的数据集，以公平的方式进行长期基准测试和比较。LLM 是在互联网的大量数据上进行训练的，这意味着 LLM 可能会使用这些现有的、通常在网上发布的评估数据集进行训练，而且可能是多次训练。这种现象被称为数据污染。这样，对 LLM 模型的评估就不再是衡量其正确回答未见问题的能力，而是简单地记忆已发布的解决方案的能力。

89. 一些研究人员还描述了未来与可用训练数据质量相关的绩效逐渐下降的情况，这种情况被称为“模型崩溃”。随着生成式人工智能系统的输出充斥互联网，未来的模型本身可能会通过早期模型创建的在线内容进行训练，其中包括早期模型的所有偏差和错误，从而导致准确性和能力的逐步下降。然而，风险并非合成数据总是有害；而是对模型生成的内容不加区分或筛选不当的使用，可能降低数据多样性并放大错误。合成数据在其生成、过滤和验证时注意质量和代表性，便能够有用。

90. 随着生成式人工智能系统与外部工具、数据库和应用程序编程接口连接，可信赖性还包括安全性。用户或第三方可能试图通过提示注入操纵系统，包括隐藏在系统检索到的文件或网页中的间接提示注入。其他风险包括敏感信息泄露、对所生成输出的不安全处理、数据或模型投毒，以及过度自主，即给与系统超出任务所需的自主权或工具权限。这些风险在代理式人工智能中更

---

19 Chelli M、Descamps J、Lavoué V、Trojani C、Azar M、Deckert M、Raynier JL、Clowez G、Boileau P、Ruetsch-Chelli C。ChatGPT 和 Bard 的幻觉率和参考准确性系统回顾：比较分析。J Med Internet Res. 2024 年 5 月 22 日；26:e53164. doi: 10.2196/53164。

20 Matthew Dahl、Varun Magesh、Mirac Suzgun、Daniel E Ho，《大型法律拟制：剖析大型语言模型中的法律幻觉》，《法律分析期刊》，第 16 卷第 1 期，2024 年，第 64-93 页，https://doi.org/10.1093/jla/laae003。

21 Varun Magesh、Faiz Surani、Matthew Dahl、Mirac Suzgun、Christopher D. Manning 和 Daniel E. Ho，《免于幻觉？评估领先人工智能法律研究工具的可靠性》，《实证法律研究杂志》第 22 期，216（2025 年）。

为突出，因为幻觉、受操纵的指令或间接提示注入攻击可引发后果严重的行动，例如发送通信、删除或修改文档、发起购买或其他交易、更改记录或调用外部服务，可能难以或无法逆转。因此，部署者应限制工具的功能、权限和自主性，并对影响大的操作要求人类确认，并有日志记录、监控和明确的备用或恢复程序。[22]

冒犯性和敏感性内容

91. 生成式人工智能模型通过海量数据、文本和图像进行训练。不可能从预训练数据中完美地过滤掉所有不恰当的、在道德伦理方面具有挑战性的内容。因此，这些模型可能会在生成的输出中复制和重新组合类似的冒犯性内容。

92. 有多种技术可以避免这种情况的发生。例如，预测的单词或短语会不断经过所谓的“护栏”进行过滤，以去除冒犯性内容（例如，“护栏”可以是黑名单上的单词）。不过，目前还不能保证完全消除此类生成输出，尤其是在使用有毒提示的情况下。

人工智能资源和基础设施

93. LLM 预训练成本高昂，而大规模利用 LLM 同样需要大型计算中心、专用硬件（如生产依赖于极少数制造商的图形处理器（GPU））和不断增加的电力能力。生成式人工智能也依赖于互联网规模的数据。结合投资能力和当地数字技能的可用性，多个论坛都在讨论高收入国家与中低收入国家之间的差距问题。[23]与此同时，人工智能技术也被用于改善发展中国家的教育、医疗保健和金融普惠等方面。[24]

94. 一些最新的技术发展可能部分减少这些障碍。更小的专用模型、模型压缩、蒸馏、量化、缓存以及端侧部署，能够降低特定应用的成本、延迟和能耗。尽管如此，对高质量数据、技术人员、计算基础设施和可靠电力的获取仍不均衡，而工具互联和多模态系统的使用增加，可能带来新的运营和安全成本。

F 人工智能与创新的当前技术格局

95. 产权组织的人工智能政策工具包根据人工智能在创新过程中发挥的作用，将人工智能发明分为四类。这种分类法很有用，因为人工智能可作为创新本身的主题，人类发明人使用的工具，

---

22 见 NIST，“人工智能风险管理框架：生成式人工智能概况”，NIST AI 600-1（2024 年）第 2.2 节和第 2.9 节（虚构内容和提示注入）；OWASP，“LLM06:2025 过度代理”，OWASP 生成式人工智能安全项目，发现过度功能、过度权限和过强自主性是导致破坏性 LLM 代理行动的根本原因；人工智能伙伴关系，“优先考虑人工智能代理中的实时故障检测”（2025 年），指出代理故障可导致未经授权的行动、不可逆的变化或现实世界中的危害，且可能发生在规划、工具使用或执行过程中；Kai Greshake 等，“非你所愿：利用间接提示注入损害现实世界 LLM 集成应用”，AISec ’23（2023 年）。

23 例如，见联合国和国际劳工组织的报告《注意人工智能鸿沟：塑造未来工作的全球视角》，2024 年 7 月 26 日，ISBN：9789211066524，https://www.ilo.org/media/581631/download。一份研究论文指出，收入水平、青年人口比例、数字基础设施、高技能可交易服务专业化、英语熟练程度和人力资本与更多的生成式人工智能采用率密切相关。见《究竟谁在使用生成式人工智能？》，政策研究工作文件 10870，世界银行集团：https://documents1.worldbank.org/curated/en/099720008192430535/pdf/IDU15f321eb5148701472d1a88813ab677be07b0.pdf。

24 例如，见《倾斜的天平：人工智能对发展中国家的双重影响》，世界银行博客：https://blogs.worldbank.org/en/digital-development/tipping-the-scales--ai-s-dual-impact-on-developing-nations。

纳入发明中的组成部分，或在未来情景中，在几乎没有或完全没有人类实质性输入情况下生成发明的系统。[25]

- 第一类是**人工智能模型**，即新的人工智能模型或算法。这可能包括新的模型架构、优化方法、训练技术，或者使人工智能系统更准确、高效或适应性更强的算法改进等。

- 第二类是**人工智能辅助发明**，即人类在发明过程中将人工智能用作工具。在这一类别中，人工智能可协助筛选选项、识别模式、选择候选方案、优化设计或测试可能的解决方案，但发明仍由人类主导。一个实例是在药物研发中利用人工智能识别或筛选候选药物。

- 第三类是**基于人工智能的发明**，即发明中囊括了人工智能，且人工智能构成产品、工艺或服务基础的一部分。在这些情况下，人工智能不仅仅是在后台使用的研究工具，作为发明过程中的辅助；而是发明如何运作的一部分。实例包括使用人工智能深度学习的翻译设备、基于人工智能的图像增强、诊断工具或其他功能依靠人工智能的系统。

- 第四类是**人工智能生成的发明**，产权组织将其描述为一种未来情景，发明由人工智能自主生成，没有实质性的人类输入。这一类别作为政策讨论的概念界限很有用，因为它有助于区分对人工智能的当前使用与可能的未来情况，即人工智能可在发明过程中发挥更独立的作用。

三、人工智能相关发明的专利保护

96. 本文件的这个部分探讨了人工智能相关发明的专利保护。如第二部分 F 节所述，“人工智能相关发明”一词在本文件中用作统称，涵盖人工智能模型、人工智能辅助发明、基于人工智能的发明以及人工智能生成的发明这一可能的未来类别。人工智能技术对专利法的影响可能不同，取决于人工智能是要求保护的发明的客体、发明如何运作的一部分，还是发明过程中使用的工具。

A. 总论

97. 专利制度应为了全社会的利益，通过给予技术生产者和技术知识使用者平衡的权利和义务，为推动技术创新以及为技术转让和传播作出贡献，这一点已得到广泛认可。各国为此提供法律框架，并颁布法律法规，法院对这些法律法规进行解释，辅以行政机关所制定的实务指导。

98. 由于专利制度对于技术持中立的态度，因此每当有新技术出现时，经常会提出一个问题，即专利制度能否继续达到其目的。半导体技术、计算机软件、信息技术和生物技术皆是如此：随着技术的发展，辩论仍在继续。因此人工智能的出现引发了同样的问题和辩论，这并不足为奇，这些问题和辩论仔细研究了目前的专利制度是否已为兼顾人工智能技术做好准备。

99. 近数十年来，包括硬件和软件在内的计算机技术被应用于协助人类在很多技术领域进行发明创造。例如，计算机辅助设计（CAD）助力了机械和电子行业的发展，生物信息学为研究人员分析和解读生物数据提供了便利，以及计算化学帮助化学家找到新化学物质。计算机还被集成进设备和装置以实现某种功能。

100. 对于计算机技术来说，与其相关的新发明可归为三类：

---

[25] 世界知识产权组织（产权组织），让创新生态系统为人工智能做好准备——知识产权政策工具包（日内瓦：产权组织，2024 年）。

(i) 改进计算机本身的计算功能的新发明；

(ii) 为实现某种功能集成了计算机的新发明（设备、装置等）；及

(iii) 在计算机的帮助下创造的可属于任何技术领域的新发明。

101. 对于人工智能技术，本文件遵循第二部分 F 节中引入的分类法，而非另行分类。从广义上讲，上述计算机技术类别对应人工智能模型、基于人工智能的发明和人工智能辅助发明，而第二部分 F 节还确定了人工智能生成的发明这一可能的未来类别。

102. 在人工智能技术现阶段的发展中，人类的指示和干预仍是创造这些发明过程中的重要部分。但是，随着人工智能技术的发展，人类干预在创造过程中的性质相比人工智能系统更多的自主执行可能会发生变化。

103. 在第二部分 F 节引入的分类法中，可通过关注人类对创造核心发明概念的贡献程度和性质，进一步理解人工智能相关的发明，特别是人工智能辅助发明以及人工智能生成的发明这一可能的未来类别。从这一角度看，可考虑以下情境：

(i) 由人类识别问题和构想解决方案，而人工智能技术仅应用于对人类的解决方案进行验证、自动化、适应或泛化；

(ii) 由人类识别问题，而解决方案的构想由人工智能技术辅助、指导或主导；

(iii) 由人类和人工智能技术共同识别问题和构思解决方案；及

(iv) 由人工智能技术识别问题和构想解决方案而没有任何人类干预。

104. 在第二种情境中，人工智能技术在发明创造过程中的相关性可以从最小化到决定性。第四种情境，即通用人工智能（AGI）或超级人工智能[26]，不是当前技术可以达到的。[27]由于定义上的模糊，尽管对于完全成熟的通用人工智能必须有什么构成尚无共识，但有观点认为窄人工智能（ANI）已经实现（例如自然语言处理、图像识别），且近期行业报告警告称，达成通用人工智能的水平可在短期内预见。但是，这一发展的可能性与传统的计算机技术之间存在显著差别。这种差别意味着就人工智能专利活动而言，人们需要提出不同性质的新问题。

105. 自人工智能技术出现以来，创新和研究人员就开始为此类发明提交专利申请并被授予专利。如在《2019 年世界知识产权组织技术趋势——人工智能》中所述，它们涉及在一系列不同人工智能申请领域[28]中的多项不同人工智能技术[29]和多种人工智能功能应用。[30]开源（或开放式创新）的方法在人工智能开发人员中也很流行。[31]关于人工智能相关发明的专利态势详细数据，请参阅所述产权组织出版物。随着人工智能在公众中的地位不断提高，科学文献中对人工智能的关注度不断增加，近年来与生成式人工智能相关的专利也呈指数级增长。[32]从专利申请量来看，最重要的生成式人工智能模型是：(i)生成对抗网络（GAN）；(ii)变分自编码器（VAE）；(iii)基于解码器

---

[26] 这意味着人工智能系统能够成功执行人类大脑可完成的任何智力任务，或者机器的假设能力远远超过人类大脑。

[27] 2019 年世界知识产权组织技术趋势——人工智能，第 19 页。

[28] 例如交通运输、电信以及生命科学和医学。

[29] 例如机器学习、模糊逻辑和逻辑编程。

[30] 例如计算机视觉、自然语言处理和语音处理。

[31] 2019 年世界知识产权组织技术趋势——人工智能，第 109 页。

[32] 世界知识产权组织（产权组织）（2024 年）。《生成式人工智能》。《专利态势报告》。日内瓦：产权组织。https://doi.org/10.34667/tind.49740

的大型语言模型（LLM）。这些专利并不集中在某一特定领域或行业，而是涉及大量关键应用领域。[33]

106. 关于人工智能技术如何影响专利法的问题尚无定论。许多国家尚未制定审查人工智能相关申请的特别程序。[34]但人工智能技术的一些特性似乎指向了专利法中可能在未来，如果不是眼下立即，受到这一新兴技术影响的领域。可供考虑的方面包括如下几点：

(i) 由于人工智能技术主要由软件实施，目前有关计算机实施发明和使用软件的发明的专利法问题对于人工智能技术可能仍然具有相关性；

(ii) 人工智能技术的认知性需要我们进一步考虑如何把这项技术融入人类创新进程，以及它对于专利制度和专利法所作出的“人造”发明假设的影响；

(iii) 在完全再现和说明在深度学习神经网络中所采用方法方面存在的固有的技术局限性使我们注意到，它们对于专利制度基本原则之一——传播新技术知识——具有潜在影响。

107. 只要专利制度的出发点仍是为推动技术创新以及转移和传播技术作出贡献，专利制度就要继续为人工智能领域也提供激励创新的举措和分享新知识的机制（除非有其他能够充分处理这些事项的法律/社会/经济工具）。在政策层面，主要的考量可以是：考虑到专利制度的目标，人工智能技术的发展是否会破坏专利制度所力求实现的平衡？如果是这样，如何恢复这种平衡？根据人工智能技术的发展对专利法和实践进行更新是否有用？专利制度的现有法律概念和人工智能的兴起之间是否存在，或将存在任何缺口？

108. 为了回答这些问题，需要了解人工智能相比传统的计算机技术具有哪些技术特性，以及需要对当前的法律和实践在目前和今后适用于人工智能技术的可能方式进行评估。本背景文件并非希望全面阐释所有问题。但以下段落提供了一些专利法问题的示例，它们在为人工智能相关发明寻求专利保护和授予专利方面可能具有相关性。“人工智能相关发明”一词指第二部分 F 节所述的类别，如上文从人类贡献角度的进一步讨论。在这个阶段，几乎没有专门处理适用于人工智能相关发明的专利法问题的正式指南。由于人工智能是一项新技术，判例法尚未得到充分发展，有为数不多的专利局发布了指南以厘清它们在该领域的做法。随着人工智能集成产品在市场中的进一步商业化发展，权利要求解释背景下的人工智能专利维权和许可也可能成为未来讨论项的一部分。一般来说，对许可协议进行谈判和解决专利争议需要复杂全面的考量。人工智能相关发明本身是否会使这一原本就很复杂的问题进一步复杂化仍有待观察。

109. 目前的专利制度建立在这样的假设基础之上，即某些激励机制会促进人类的创造活动。从高级别政策的角度来看，人工智能技术发展所展示出的潜力对专利制度的激励理论构成了一个法律哲学问题。人工智能机器能够以自主的方式全面处理各种不同数据（不仅是科技数据，还有个人和行为数据以及社会和法律数据）、识别问题、通过新发明解决问题以及为市场生产新产品以满足人类需求，尽管这还只是科一种遐想，但一旦实现，上述问题将尤为突出。尽管这可能是一个在智力上耐人寻味的问题，但它远超出了本文件的范围。

110. 人工智能的最新进展，尤其是在深度学习和生成模型等领域的进展，使得人工智能系统越来越先进，能够在极少人工干预的情况下完成复杂的任务。这些发展引发了有关人工智能生成发

---

[33] 例如，软件、生命科学、出版、商业解决方案、工业、交通、安全和电信。以坦桑尼亚为实例，《专利法》第217章未明确提及人工智能或生物技术等新兴技术，但新颖性、创造性和工业实用性的一般可专利性要求适用于所有类型的发明，包括新兴领域的发明—见坦桑尼亚答复C.9304的评论意见。

[34] 例如俄罗斯联邦，见俄罗斯联邦答复C.9199的评论意见。

明对现行专利制度影响的重要讨论。[35]在现行专利法下人工智能能否被视为发明人的争论愈发引人关注，一些司法管辖区还就此发布了指导意见。

111. 此外，人工智能与各行各业的快速融合也凸显出为了专利申请过程中的法律确定性、一致性和公平性，有必要就人工智能相关发明的可专利性制定更明确的指南和标准。人工智能能够自主学习和演化这一特性对新颖性、创造性和充分公开等传统可专利性要求提出了挑战。一些人工智能模型的黑匣子性质，尤其是在深度学习方面，使得发明说明书必须清晰及可再现的要求变得更为复杂。

B. 可授权主题

112. 一般来说，所有技术领域的任何发明，无论是产品还是方法，都可授予专利，条件是发明符合所有法律要求，包括发明不属于不可授权主题的要求。对于“发明”一词在国际层面没有强制性定义，国内法对不可授权主题的范围进行界定，使其与该国是缔约方的国际条约相一致。因此，各国对于可授权主题范围的界定有所不同。[36]很多国家把数学方法、机制、从事智力活动的规则和方法、商业规则和方法以及计算机程序排除在可授权主题之外。其中一些国家作出了澄清，即这些主题仅在专利申请与该主题本身相关的范围内被排除在可授权主题之外。在一个管辖区[37]，判例法规定，针对自然规律、自然现象和抽象概念的权利要求排除在专利保护之外。在另一个管辖区，其专利法[38]规定“发明”是“技术概念利用自然规律实现的高度先进的创造”，并且产品发明的类别中包括计算机程序和由等同于计算机程序的电子计算机处理的任何其他信息。[39]

113. 使用第二部分 F 节引入的分类法，本节主要涉及人工智能模型和基于人工智能的发明的可专利客体问题，而人工智能辅助的发明将在相关部分另行讨论。除了对运行人工智能功能的硬件组件所作的改进外，与人工智能技术和人工智能功能应用有关的发明主要指向软件。因此，许多成员国认为人工智能相关发明是计算机实施发明的一个子集。[40]与传统计算机技术的情况一样，人工智能应用也可用于非技术领域，如金融、保险、电子商务等。此外，机器学习基于计算模型和算法来分类、聚类、回归和降维，这可能被视作数学方法。另外，尽管不能否认训练数据对于机器学习绩效的重要性，但只是属于信息的数据本身不是可授予专利权的发明。

114. 计算机实施发明（CII）或软件实施发明的专利适格性已经是难以在适格主题和不适格主题之间作出明确区分的领域之一。例如在很多国家，要求保护发明的“技术性”被认为是判断专利适格性的重要因素。在这些国家建立了判例法和主管局实践，以厘清“技术问题”、“技术手段”、“技术效果”和“技术目的”等概念。

115. 人工智能通常被视为计算机科学的一个分支，因为涉及对计算和信息的研究，包含算法数据结构、编程语言、计算机体系结构和软件工程，其中包括以高效方式表示、存储、处理和传输计算输出及信息的方法和手段。[41]但是，关于人工智能模型和算法是否能够轻易被认定为 CII，存

---

35 例如，见澳大利亚和巴西答复 C.9199 的评论意见。

36 见“国家/地区专利法的若干方面——从可授权主题中排除的主题”：https://www.wipo.int/scp/en/annex_ii.html。

37 美利坚合众国。

38 日本专利法 2(1)和(4)条。

39 更多有关从可授权主题中排除的主题和计算机实施发明的专利适格性的信息可参见 SCP/13/3 和 SCP/15/3（关于计算机程序不属于可授权主题的内容请特别参见 SCP/15/3 附件二）。

40 例如，见智利、立陶宛和葡萄牙答复 C.9199 的评论意见。

41 “人工智能是计算机科学吗”，CRLN 团队（2025 年），可参见：https://www.clrn.org/is-ai-computer-science/

在重大争议。[42]人工智能的统计性、概率性和非确定性带来了新的挑战和复杂性，出于多种原因，现有的 CII 框架可能无法完全涵盖。计算机程序由编程和基于指令的输出组成，而人工智能模型和机器学习系统则是通过基于训练的内部逻辑运行，而非编程。因此，不同于涉及明确且确定性功能步骤、信息输入与输出之间关系清晰的传统计算机程序，经过训练的神经网络由数值参数定义，它们与网络行为之间的关系可能仍不明确，即便经过人类充分审查以追溯人工智能系统的思维过程或其决策原因（即通常称为“黑箱”问题）。[43]因此，CII 与人工智能发明之间的差异，在客体资格的背景下，对专利法和政策提出了一些根本性的理论问题。

116. 尽管相关理论问题尚未得到广泛讨论，在学术文献中除外，但至少有一个司法管辖区的司法发展已提出此类关切，并提供了指导性的政策见解。在 Emotional Perception AI 诉专利、外观设计和商标局局长案中，联合王国高等法院认定，不能假设人工智能模型是计算机程序，因而属于不可授予专利的客体，就基于现行 CII 框架的适用将依赖人工神经网络（ANN）的人工智能发明排除在外。[44]在该案背景下的讨论表明，人工智能的使用（即用于媒体文件推荐的 ANN）并不完全符合“计算机程序”的传统概念。该案中至少两项关键观察结果突出了人工智能模型客体资格的理论问题。第一，法院将人类编写的传统程序指令与经过训练的 ANN 的运行区分开来，后者的运行参数通过学习产生，并非作为人类给出的代码执行。因此得出结论，在该案的具体情况下，经过训练的模拟 ANN 能够与底层软件区分开来，就排除而言，其本身并不构成“计算机程序”。第二，案件表明，在对经过自训练的 ANN 适用计算机程序排除规定时，需要仔细确定相关计算机、相关程序、实际要求保护的内容以及要求保护的发明是否产生了相关技术效果（如有必要）。尽管在启动或实现训练的过程中涉及计算机程序，但法院认为这仅仅是权利要求的次要部分，不代表要求保护的发明。[45]尽管该案后来在上诉中被推翻（见下文第 131 段），但法院提出的意见为理解 CII 与人工智能发明之间的不同点提供了分析框架。

117. 对于“灰色地带”的人工智能技术，CII 框架还带来了进一步的困难，最突出的是人工智能模拟。人工智能模拟利用数据驱动的算法来检测模式、生成输出并优化流程，而无需显式编程。需要模拟来提供训练人工智能模型的真实世界情景。例如，对于在工厂车间不同用例中使用的短距离移动型人工智能设备，模拟真实世界环境很重要，用于训练人工智能模型以及基于为此目的使用人工智能模型创建的合成数据进一步测试，从而补充和优化模型的部署。作为通过学习的参数将输入转换为输出的数学函数，人工智能模拟主要产生信息性输出（即预测、分类或推荐），不一定对硬件造成具体改变。因此，专利可能因输出构成基于抽象数学概念/方法的信息性内容而被驳回。欧专局扩大上诉委员会在 G1/19“行人模拟”（2021）案中，将现有判例法应用于计算机实施的模拟问题，其中涉及通过计算机程序实施的方法以及用于模拟行人群体在火车站等环境中移动的装置以进行人群管理，得出结论认为下游技术效果必须与物理现实相关。[46]因此，用于辅助其他解决方案的设计或验证的人工智能模拟可保留技术特征，前提是所主张的技术效果明确确立且对权利要求有适当限定。

---

42 产权组织人工智能与知识产权工具包，第 13 页。

43 见上文第 80 段和第 81 段。另见“人工智能的神秘‘黑箱’问题解析”，密歇根大学迪尔伯恩分校（2023 年），可访问：https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained

44 [2023] EWHC 2948 (Ch)。案件涉及将人工神经网络（ANN）用于提供改进的媒体文件推荐系统，并基于人类认知和情感向最终用户发送文件和消息（无论音乐流派及传统分类如何）这一用途的可专利性问题。

45 同上，第 32、54、56-61、63 和 76-78 段。

46 2021 年 3 月 10 日 G 0001/19（行人模拟）案，欧专局扩大上诉委员会

118. 以下对不同司法管辖区的调查表明，没有司法管辖区对人工智能发明的专利适格性专门修正专利立法。但是，在关于客体适格性改革的更广泛讨论背景下，特定司法管辖区已出现立法建议。此类建议若通过，可能会通过限制司法制定的客体排除，并以一份法定不适格类别的封闭清单代替，对人工智能发明的处理产生实质性影响。根据此类提案，适格性评估将把要求保护的发明作为整体考虑，如本领域普通技术人员的理解，重点关注发明是独立于人类活动存在于自然界中，还是仅存在于人类思维之中。[47]

119. 至于适用于人工智能相关发明的专利适格性要求，一些专利局发布了与人工智能相关发明有关的指南，具体如下。

120. 在澳大利亚，澳大利亚知识产权局认为，就客体适格性而言，人工智能相关发明通常是计算机实施发明的一个子集。尽管澳大利亚法院尚未具体审理过包括或使用人工智能的发明可专利性，但关于计算机实施发明的最新进展与人工智能相关申请具有相关性。在 Aristocrat Technologies Australia Pty Ltd 诉专利专员[2025] FCAFC 131 案中，联邦法院全席法庭确认，应从实质上对计算机实施的发明做出界定，相关的问题在于要求保护的客体仅仅是在计算机上操作的抽象想法，还是在计算机上实施、产生人为状态和实用结果的抽象想法。在专利专员诉 Aristocrat Technologies Australia Pty Ltd [2026] HCADisp 15 案中，高等法院驳回了特别上诉许可申请。

121. 继 Aristocrat 案之后，澳大利亚知识产权局更新了其《专利实务和程序手册》。该手册现规定，计算机实施的发明若经恰当界定，涉及在计算机上实施抽象想法以产生人为状态和实用结果，则可能构成可专利客体，并且提供了一个决策框架，考虑权利要求的物理要素、计算机化产品或系统的运行方式是否得到改进或改变、相关技术效果是否发生在计算机化产品或系统之外，以及技术问题是否通过该计算机化产品或系统得到解决等。根据这些原则，如果人工智能或计算机实施，经恰当界定，不仅仅是对通用计算机技术上抽象想法的操作，而是为此类可专利的实施做出贡献，则人工智能相关权利要求可被认定为具有可专利性。[48]

122. 在巴西，国家工业产权局（INPI）于 2020 年更新了计算机实施发明（CII）指南，以处理与人工智能相关的专利申请。[49]专利适格性的确定取决于要求保护的主题是否落入 INPI 审查指南中的可专利性排除范围。修订后的指南还强调，人工智能技术，包括机器学习和深度学习工具，在应用于解决技术问题时，可被视为发明。[50]在此，如果人工智能模型或算法脱离了某一特定技术领域的申请，则被视为数学方法或算法，因此不能被授予专利。相比之下，可授权的主题包括与人工智能核心技术变化有关的发明，如改变训练过程或开发新的神经网络架构，前提是如果考虑到所要解决的具体技术问题，这种变化是合理的，此外还要被纳入某一技术领域并产生技术效果。在巴西，神经网络架构单独来看被视为数学方法。关于基于硬件的人工智能系统，为确保专利适

---

47 国会研究服务处，“专利适格的客体改革：概述”（2026 年）。可访问：https://www.congress.gov/crs_external_products/IF/PDF/IF12563/IF12563.4.pdf

48 见澳大利亚答复 C.9199 的评论意见；Aristocrat Technologies Australia Pty Ltd 诉专利专员[2025] FCAFC 131；专利专员诉 Aristocrat Technologies Australia Pty Ltd [2026] HCADisp 15；澳大利亚知识产权局，“关于计算机实施发明的专利专员指导更新”，官方通告，2025 年 10 月；澳大利亚知识产权局，“更新：Aristocrat Technologies Australia Pty Ltd 诉专利专员案”，官方通告，2026 年 2 月 6 日；澳大利亚知识产权局，《专利实务和程序手册》，第 5.6.8.6 节，“计算机实施的发明、单纯方案和商业方法”，2026 年 3 月 16 日更新；澳大利亚知识产权局，“澳大利亚知识产权局专利手册更新——2026 年 3 月”。

49 2020 年决议 INPI/PR 第 411 号。

50 见巴西答复 C.9199 的评论意见。

格性，必须说明硬件实施的具体特征和细节，仅提及可能的硬件解释是不够的。[51]2025 年，INPI 就《人工智能相关专利申请的审查指南》草案启动了第 03/2025 号公开意见征求；该草案将人工智能相关发明作为计算机实施的发明子群对待，并保留了可授予专利的人工智能相关发明解决技术问题、提供技术解决方案并产生技术效果的要求。[52]

123. 在中国，《专利审查指南》已于 2024 年 1 月 20 日起修订生效，其中涉及含人工智能和大数据的发明专利申请的审查等事项。根据该指南，若要求保护的解决方案作为整体考虑，利用技术手段解决技术问题并产生技术效果，则人工智能相关发明可构成可专利客体。例如，若算法与计算机系统的内部结构有特定技术关联，并提高了硬件计算效率或改进执行结果，则该人工智能或大数据算法的改进可构成技术解决方案。2024 年 12 月 31 日，国知局发布了《人工智能相关发明专利申请指引（试行）》，其中提供了更多针对人工智能的具体指导和示例。该指引明确指出，人工智能相关权利要求不得仅属于思维活动的规则和方法，还必须满足技术方案的要求；同时进一步阐明，如果权利要求的实质仍为缺乏技术问题或技术效果的抽象数学算法或模型训练方法，则仅增加人工智能相关技术特征或通用计算机硬件是不够的。《专利审查指南》随后由第 84 号令进一步修正，自 2026 年 1 月 1 日起生效，其中包括增加人工智能和大数据相关条款，涉及包含算法特征或商业规则和方法的权利要求；并规定，涉及违反法律、社会公德或公共利益的数据收集、标签管理、规则设定或推荐决定的申请，不得授予专利权。[53]

124. 在法国，国家工业产权局（INPI）在其审查指南中，对决定可专利客体适用基于《知识产权法典》第 L. 611-10 条的分析框架。特定客体不视为发明，特别是数学方法、计算机程序和商业方法，但仅限于以该形式要求保护时。对于使用人工智能的发明，INPI 不适用单独的制度：人工智能只要是基于计算，就被视为由计算机实现的数学方法，若以此提出权利要求，则不得视为发明。若使用人工智能的发明要具有可专利性，该人工智能必须通过非通用技术手段或通过对测量技术数据的处理，为技术问题提供技术解决方案，对发明的技术特征做出贡献。视为技术性的实例包括图像或视频的处理、识别或分类，机器人技术中的实时控制或监测和控制过程，利用机器学习对 IP 流量进行分类以改善流量管理，以及在心脏监测设备中使用神经网络开展预测分析并检测心律不齐。[54]

125. 到目前为止，德国还没有确立专门针对人工智能相关发明可专利性的判例法。不过，由于这些发明在概念上类似于计算机实施发明，因此，通常采用德国联邦法院（BGH）根据德国专利法第 1、3 和 4 条制定的程序相关发明三步审查法来处理专利适格性问题。[55]

126. 在印度，2025 年《计算机相关发明审查指南》（CRI 指南）中载有关于人工智能、机器学习和深度学习相关发明的专门章节。指南指出，此类技术根本上建立在数学模型、学习算法、计算模型和抽象原理之上。抽象想法、数学公式或理论上的人工智能/机器学习/深度学习构造因缺乏

---

51 同上。

52 INPI，Consulta Pública INPI/PR No. 03/2025，“Minuta das Diretrizes de Exame de Pedidos de Patente relacionados à Inteligência Artificial”，咨询期为 2025 年 8 月 19 日至 10 月 19 日；INPI，Minuta das Diretrizes de Exame de Pedidos de Patente relacionados à Inteligência Artificial，2025 年 8 月，第 1.1、1.3－1.4 和 2.1－2.2 节。

53 国知局，“《专利审查指南》（2023）修改解读（四）——涉及计算机程序的发明专利申请审查”，2024 年 1 月 18 日；国知局，“人工智能相关发明专利申请指引（试行）”，2024 年 12 月 31 日；国知局，人工智能相关发明专利申请指引（试行）常见问题解答，2024 年 12 月 31 日；国知局，关于修改《专利审查指南》的决定（局令第 84 号），2025 年 11 月 13 日，自 2026 年 1 月 1 日起生效。

54 见法国答复 C.9304 的评论意见。

55 见德国答复 C.9199 的评论意见。

实际应用而不具有可专利性，不过当人工智能/机器学习/深度学习创新将抽象原理转化为现实世界的有形应用，并通过技术手段为问题提供可实施的技术解决方案时，则可具备可专利性。该指南还确定了技术实施，例如为具体技术任务提高性能或降低计算成本的机器学习模型高效训练，或具有创造性的神经网络架构，作为可避免被当做计算机程序本身排除的示例。[56]

127. 自 2019 年德里高等法院在 Ferid Allani 诉印度联邦案中做出允许计算机相关发明在证明“技术效果”时具有可专利性的裁决以来，印度各法院随后也根据第 3(k)条处理人工智能/机器学习相关的专利适格性。在 Vandana Parvez 博士诉专利主管官员案中，马德拉斯高等法院撤销了对一件机器学习相关在线广告和个性化申请的驳回，指出其初步证明的技术进步和技术效果。相反，在 Caleb Suresh Motupalli 诉专利主管专员案中，马德拉斯高等法院维持了对一项人工智能、增强现实和混合现实相关申请的驳回，认为要求保护的发明缺乏实施实例和技术效果，不满足第 3(k)条。[57]

128. 日本特许厅发布的专利和实用新型审查指南中也包含与人工智能发明有关的示例。[58]这些示例说明了判定人工智能发明是否属于利用自然规律的技术想法创造标准，这是获得专利适格性的关键。在大韩民国，知识产权部（MOIP）编拟了一份详细的人工智能领域审查指南。[59]与其他主管局类似，指南指出，人工智能相关发明的专利适格性标准原则上与计算机软件相关发明的标准相同。一般来说，人工智能发明在以下情况下符合授予专利的条件：(i)信息处理是通过“软件和硬件的组合”进行实施；(ii)权利要求不涉及“人类智力活动或离线活动”。MOIP 编拟的审查做法指南还包括与人工智能发明有关的示例。

129. 新加坡同样为人工智能相关发明的专利适格性判断提供了具体指导。[60]根据审查指南，人工智能和机器学习方法，如神经网络、支持向量机、判别分析、决策树和 k-means，被视为数学方法，其本身不被视为发明。[61]仅仅使用传统计算机硬件实施人工智能方法不太可能符合适格性标准，除非实际贡献超出了基础数学方法的范围。不过，应用人工智能解决具体问题的人工智能相关发明，如识别人类语言或图像，可能被视为符合专利适格性。通过在问题与数学方法的步骤之间建立充分的联系，权利要求应在功能上明确或隐含地限制在解决具体问题上。[62]例如，具体说明一连串数学步骤的输入和输出如何与问题相关，可确保该方法与解决问题之间存在因果关系。此外，提及计算机实施或使用计算机硬件的要求保护的人工智能方法必须证明它们解决了一个具体问题。如果申请仅仅涉及使用传统硬件来实施机器学习方法，那么除非硬件交互实质性地解决了一个具体问题，否则就不太可能符合专利适格性。

130. 此外，2024 年 10 月，新加坡知识产权局（IPOS）发布了《人工智能相关专利申请审查补充指南》。这份资料通过进一步明确人工智能发明的可专利性，对现有专利审查指南进行了补充。

---

56 专利、外观设计和商标总署，2025 年《计算机相关发明（CRI）审查指南》，第 4.5 节和第 5.1-5.2 节

57 马德拉斯高等法院，Vandana Parvez 博士诉专利主管官员案，2023 年 CMA(PT)第 33 号（2024 年 2 月 23 日）；马德拉斯高等法院，Caleb Suresh Motupalli 诉专利主管官员案，2024 年 C.M.A.(PT)第 2 号（2025 年 1 月 29 日）

58 专利和实用新型审查手册附件 A 和 B。在专利适格性方面，讨论的示例包括：指向仅仅是信息呈现的数据的权利要求；能够进行信息处理的数据结构，这种处理可以在语音交互系统中进行；用于分析住宿场所声誉的经训练的模型。

59 “按技术领域分列的审查做法指南”，知识产权部，可访问：https://www.kipo.go.kr/upload/en/download/EXAMINATION_PRACTICE_GUIDELINES_BY_TECHNOLOGY_FIELD.pdf；见大韩民国答复 C.9304 的评论意见。

60 IPOS 专利申请审查指南，第 8.22-8.27 段。

61 同上。

62 同上。

其中阐述了当前的审查做法，并包括了假设性示例，说明评估人工智能领域发明可专利客体的方式。[63]

131. 在联合王国，法院曾在 Emotional Perception AI Limited 诉专利、外观设计和商标局局长案中审议了人工神经网络（ANN）的客体适格性。高等法院曾认定，将 ANN 用于媒体文件推荐在计算机程序排除规定的范围之外，因为人工智能驱动的 ANN 通常不视为计算机程序（见上文第 116 段的讨论）。上诉法院则认定，ANN 无论是通过硬件还是软件实施，均属于计算机程序排除范围内，且要求保护的推荐系统并未产生相关技术效果。2026 年 2 月，联合王国最高法院允许上诉。最高法院认定，ANN 属于计算机程序排除规定范围内的计算机程序，但要求保护的客体并非“这样的”计算机程序，因为主张保护的方法涉及技术手段，包括实施 ANN 的硬件，以及在相关权利要求中的数据库、通信网络和用户设备。最高法院拒绝了 Aerotel 案的方式，并采纳了欧专局扩大委员会在 G 1/19 案中对《欧洲专利公约》（EPC）第 52 条的相同解释，即“发明”要求是一个低门槛，与新颖性、创造性和工业适用性分开评估。申请被发回联合王国知识产权局，以适用“中间步骤”和可专利性的其余要求。[64]继判决做出后，联合王国知识产权局撤回了其人工智能审查指南和配套情境；联合王国知识产权局《专利实践手册》指出，其指导方针可能会发生变更，有待对最高法院判决的审阅。[65]

132. 在美利坚合众国，美国专利商标局（美国专商局）发布的 2019 年专利主题适格性指南修订版包含一个示例，具体讨论了一种训练神经网络进行面部识别的计算机实施方法的专利适格性，该方法包括一系列训练步骤。[66]随后，美国专商局于 2024 年更新了人工智能发明专利适格性指南。更新后指南的核心是适用最高法院对于“Mayo、Myriad 和 Alice”案判决的框架。美国专商局主题适格性分析的步骤 1 处理的是要求保护的发明是否是一种新的和有用的方法、机器、产品、物质组成或其改进（即法定主题）。如果一项发明不属于方法、机器、产品或物质组成，则该发明在法律上不可授予专利。步骤 2A 第一项涉及判断一项权利要求是否指向可专利性的法定例外，如抽象概念、自然法则或自然现象。如果权利要求没有记载法定例外，则认为符合专利适格性，专利适格性分析就此结束。如果确定存在法定例外，步骤 2A 第二项就会评估权利要求项目单独或组合在一起是否明显超出法定例外，从而使其转变成为符合专利适格性的申请。如果权利要求“指向”法定例外，则分析进入步骤 2B，评估要求保护的附加项目是否明显超出所记载的法定例外本身。步骤 2B 包括考虑附加项目是否属于充分了解、常规、传统的活动。基于一个或多个法定考虑因素（例如，认为附加限制是无关紧要的额外解决活动或为适用例外而提供的更多指示），一项权利要求可能会被认定为未明显超出（因此不适格）。相比之下，如果美国专商局的工作人员在步骤 2B 中判断附加项目确实明显超出法定例外，则权利要求符合专利适格性。

133. 关于人工智能发明，更新后的指南特别对上述测试的两个方面作了进一步澄清：(1)在步骤 2A 第一项中对权利要求是否记载了抽象概念的评估；(2)在步骤 2A 第二项中对改进考虑因素的评估。更新后的指南提供了大量评注和假设性示例，说明如何将测试中的这些步骤适用于人工智能

---

63 见新加坡答复 C.9304 的评论意见。

64 Emotional Perception AI Limited 诉专利、外观设计和商标局局长[2026] UKSC 3，联合王国最高法院，2026 年 2 月 11 日判决。

65 UKIPO，《人工智能相关专利申请审查指南》，于 2026 年 2 月 11 日继最高法院对 Emotional Perception AI Limited 诉专利、外观设计和商标局局长[2026] UKSC 3 做出判决后撤回；UKIPO，《人工智能专利申请审查指南的应用场景》，于 2026 年 2 月 11 日撤回；UKIPO，《专利实践手册》第 1 节和第 3 节，指出指南可能修改，有待对该判决的审查。

66 2019 年修订的专利主题适格性指南，例 39。

相关技术。[67]虽然新指南强调分析主题适格性的“Alice/Mayo 测试”并未发生变化，但新指南引入了三个新的或是通过或是未通过测试的示例。这些示例涉及“使用神经网络进行异常检测的人工智能”、“语音信号分析人工智能”和“个性化医疗人工智能”。[68]美国进一步指出，2024 年 7 月的指南以现有指南为基础，进一步明确美国专商局和申请人应如何评价人工智能技术相关发明的专利申请和专利中权利要求的客体适格性，该指南并未将人工智能发明与其他发明区别对待。[69]

134. 在欧专局，人工智能和机器学习在《审查指南》关于数学方法的指导下提及。关于人工智能和机器学习的专门小节在2018年版的指南中引入，并在2025年版的指南中得到更新。当前《指南》指出，人工智能和机器学习基于计算模型和算法，例如人工神经网络、遗传算法、支持向量机、k-均值算法、核回归和判别分析，它们本身具有抽象的数学性质，无论是否能够利用训练数据进行训练。但是，人工智能或机器学习的使用本身并不导致发明丧失可专利性。如果针对人工智能或机器学习相关发明的权利要求涉及使用技术手段（如计算机）的方法，或涉及一种装置，则其客体作为整体具有技术特征，不根据 EPC 第 52 条第(2)款和第(3)款排除在可专利性之外。在这种情况下，若计算模型或算法为技术问题的技术解决方案做出贡献，例如通过应用于技术领域或适应特定的技术实施，它们本身便为发明的技术特征做出了贡献。[70]

135. 欧专局指南给出了技术应用的示例，包括在心脏监测装置中使用神经网络来识别心律不齐，以及基于低级特征对数字图像、视频、音频或语音信号进行分类。相反，仅根据文本内容对文本文件进行分类，或者在未说明所得分类结果的技术用途情况下对抽象数据记录分类，本身不视为技术性目的。[71]

136. 2025 年 4 月，在 Recentive Analytics, Inc. 诉 Fox Corp. 一案中，美国联邦巡回上诉法院重申了这一方式，并确认通常用于评估专利客体适格性的两要件检验标准也应适用于人工智能发明。2025 年 8 月，美国专商局向特定技术中心的专利审查员发布了一份备忘录，就评估软件相关发明（包括人工智能和机器学习技术）的专利客体适格性提供了有针对性的提醒；该备忘录强化了美国专商局的现行做法，并澄清了在应用专利适格性分析时的常见问题。[72]在 2025 年 9 月 Ex parte Desjardins 案的决定中，美国专商局认定一组人工智能机器学习权利要求符合专利客体资格，发现尽管这些权利要求涉及以数学算法形式呈现的抽象想法，但记载了将该抽象想法整合到实际应用中的额外要素。具体而言，出于判定客体适格性之目的，认为权利要求充分主张了人工智能机器学习模型运行如何改进，而不仅仅是底层的数学计算；因此，Desjardins 的决定确认，人工智能机器学习发明不会仅因其记载了数学算法就一概排除在可专利性之外。[73]

137. 2025 年 12 月，美国专商局还发布了两份备忘录，就客体适格性声明（SMED）的使用提供了进一步的澄清。根据美国专商局的规定，申请人可提交 SMED 以提供与适格性审查相关的事实证据，例如技术改进的证据、提交时的现有技术，或证明司法例外如何融入实际应用的信息。第一份备

---

67 2024 年关于专利主题适格性（包括人工智能）的指南更新：https://www.federalregister.gov/documents/2024/07/17/2024-15377/2024-guidance-update-on-patent-subject-matter-eligibility-including-on-artificial-intelligence。

68 2024 年 7 月主题适格性示例：https://www.uspto.gov/sites/default/files/documents/2024-AI-SMEUpdateExamples47-49.pdf。

69 见美利坚合众国答复 C.9304 的评论意见。

70 《欧洲专利局审查指南》，G 部分，第二章，第 3.3 和 3.3.1 节，“人工智能和机器学习”；欧专局，2025 年修订版中修订的章节列表，G 部分，G-II，3.3.1，“对人工智能和机器学习定义的澄清”；欧专局，《计算机实施的发明索引》，其中指出 G-II 第 3.3.1 节被引入 2018 年版的指南，并在 2025 年版的指南中更新。

71 同上。

72 见美利坚合众国答复 C.9304 的评论意见。

73 见美利坚合众国答复 C.9304 的评论意见。

忘录面向审查队伍，解释 SMED 为自愿选项，并明确当 SMED 妥善提交时，审查员必须将其视为证据记录的一部分，并使用证据优势标准加以评价；第二份备忘录面向申请人和从业人员，指出 SMED 应提供与要求保护的发明相关的客观证据，且不应用于补充初始公开内容。

138. 上述两份备忘录共同强调，SMED 并不改变美国专商局的现行程序，而是阐明申请人可如何提供与适格性分析相关的事实证据，以及审查员应如何依据长期做法评估此类提交材料。[74]

139. 在菲律宾，现行的《知识产权法典》并未对人工智能在发明或外观设计生成中的使用程度规定禁止，且此类使用本身不得作为认定发明不具可专利性的依据。但是，在版权法、反不正当竞争法和侵权法等其他相关法律下，过度使用人工智能可能会产生相应的法律后果。[75]

140. 关于人工智能辅助发明，对可专利客体的考量取决于最终要求保护的发明的性质及其权利要求的撰写方式。例如，在植物被排除在可授权主题之外的国家，对在人工智能工具协助下创造出来的创新型新植物进行了界定的专利权利要求不具有可专利性。

141. 在人工智能模型方面，针对训练数据、数据分类、训练数据的方法、数据结构以及经训练的模型，已出现关于专利适格性的不同关切。针对训练数据“本身”的权利要求通常仅被视为信息的呈现，不符合可专利客体的资格，尽管训练数据在人工智能模型的表现上有至关重要的作用。日本特许厅的指南中提供的示例清楚地说明，针对训练数据本身的权利要求不具备可专利性。作为日本特许厅人工智能专利指南示例引用的示例 3 的权利要求 1 涉及“苹果的含糖量数据”，该权利要求针对的是数据本身。因此，根据日本特许厅的规定，“权利要求 1 在信息呈现（即呈现本身、呈现手段或呈现方法）上不具备技术特征，其特征仅在于信息的内容，且主要目标是呈现信息”。同样，该指南示例 5 的权利要求 1 称，将训练数据用于“训练判别器”是判别器机器学习中非常常规的操作。根据日本特许厅，这样的权利要求“没有为判别器的学习手段或方法带来任何技术特征”，且被视为“仅以由训练数据的图像构成的信息内容为特征”，而训练数据通过结合多个数据输入生成。[76]欧专局的指南和判例法也明确强调，针对单纯的数据分类或此类数据的处理（包括来源于技术手段的数据）的专利权利要求不具备可专利性，因为分类结果与技术目的或技术应用之间不存在关联。[77]

142. 生成训练数据的方法，或基于对象的技术特性具体开展信息处理的方法，在认定软件和硬件共同运作的情况下，可能满足适格性要求。日本特许厅人工智能专利指南中作为示例引述的示例 5 的权利要求 3 允许针对人体放射学图像的权利要求，当此类图像与手术器械图像结合以生成新图像用作训练数据时。此类训练数据可通过软件利用硬件资源执行定义好的信息处理操作生成。[78]同样，根据美国专商局的人工智能专利客体适格性指南中的示例 47 权利要求 2，该权利要求涉及人工神经网络（ANN）的训练和使用，被认定不构成符合专利适格性的客体。该要求保护的方法涉及接收和离散化训练数据、使用反向传播和梯度下降算法训练 ANN，以及利用经训练的 ANN 检测和分析数据集中的异常。相关步骤被认为涵盖了数学概念和思维过程，而使用通用计算机和经训练的 ANN 并未将这些抽象想法整合到实际应用中，也未提供创造性概念。因此，认为该权利要求不符合专利保护资格。[79]

---

74 见美利坚合众国答复 C.9304 的意评论见。
75 见菲律宾答复 C.9304 的评论意见。
76 日本特许厅《专利和实用新型审查手册》附件 A。
77 T 1784/06（分类方法/COMPTEL），2012 年 9 月 21 日。
78 日本特许厅“关于专利适格性和工业适用性的附件 A 案例”中的示例 5
79 美国专商局，2024 年关于专利客体适格性（包括人工智能）的指导更新（2024 年 7 月客体适格性示例）。

143. 当数据结构在功能上等同于满足软件-硬件交互要求的程序时，该数据结构可能符合资格。日本特许厅人工智能专利指南中作为示例引用的3D打印数据示例（案例2-15）指出，在特定情况下，如果结构化数据的结构决定了计算机或其他技术系统开展的处理程序，则该数据可被视为符合专利资格的客体。举例而言，对于三维物体的每一层，3D 打印数据可能包含模型材料数据和支撑材料数据的指针。这些指针可使3D打印机的控制单元从存储单元中检索相关数据，并在打印过程中执行相应的操作。在这样的情况下，数据结构定义了3D打印机控制单元执行的信息处理，并用于控制设备的运行。因此，结构化数据可具有相当于计算机程序的特征，只要其结构决定了技术系统执行的处理。根据适用的国家法律，此类结构化数据因此可被视为符合专利适格性的客体。[80]

144. 被认定为程序的经训练模型可能满足软件-硬件交互要求。美国专商局在人工智能专利适格客体背景下作为示例引用的案例 39，涉及训练面部检测神经网络的计算机实施的方法。该方法通过对现有面部图像应用镜像、旋转、平滑或降低对比度等变换，并将所得图像与面部图像和非面部图像结合生成训练数据，并形成训练数据集。在初始训练阶段之后，将错误分类为面部图像的非面部图像添加到训练数据集中，再次训练神经网络。如此，要求保护的方法被视为符合专利适格性的客体，因为其构成了一种过程且本身未涉及司法例外（即抽象想法），尽管该方法的特定方面可能基于数学概念。[81]日本特许厅的示例（案例 2-14）涉及经训练的模型，当作为促使计算机执行具体信息处理的程序运行时，其权利要求被视为符合专利适格性的客体。在该案例中，经训练的神经网络模型使用训练权重和涉及特定词汇频率的输入数据，计算并输出与住宿设施声誉相关的量化值。由于该经训练的模型定义了通过软件与硬件资源交互执行的具体处理，并由此促使计算机执行特定功能，因此被认定构成发明，尽管要求保护的客体被描述为模型而非程序。[82]

145. 上述专门针对权利要求涉及训练数据、数据结构和经训练模型的人工智能专利适格性的示例表明，对待专利适格性的不同方式在一些关键要素上趋于一致。人工智能相关数据或模型的特征本身并不能决定其是否有资格获得专利保护。需考虑其在要求保护的发明框架下的功能性或技术性贡献。仅作为信息主张或主要以信息内容定义的训练数据或数据分类，可能不构成符合专利资格的客体。相反，生成或处理训练数据、结构化数据和经训练的模型的方法，若为技术目的或应用做出贡献，决定技术系统的运行或处理，或通过软件与硬件的交互实施特定信息处理，则可能具备适格性。同时，若要求保护的客体仍针对抽象想法，没有进一步的技术实施或应用，则仅使用计算机、数学方法、ANN 或其他人工智能技术可能不够。相应地，人工智能背景下的专利适格性可能较少取决于要求保护的客体是否被标记为数据、数据结构、训练方法或经训练的模型，而更多取决于要求保护的客体的性质和功能、其与技术系统或目的的关系，以及国家或地区专利法中适用的要求。

C. 新颖性和创造性

146. 据说创造性分析是专利性标准中最难评价的要求。[83]在被驳回的专利申请中，很多申请的驳回理由是缺乏创造性。当专利的有效性遭到第三方质疑时，它们经常把与创造性要求不符作为其主张的依据。人工智能领域的专利申请和专利似乎面临着同样的问题。已出现了一些问题，即对

---

80 日本特许厅《专利和实用新型审查手册》附件 A[案例 2-15]3D 打印数据。

81 美国专商局《客体适格性示例》示例 39：抽象想法。

82 日本特许厅《专利和实用新型审查手册》附件 B。

83 更多关于创造性要求如何在不同国家实施的信息见 SCP/22/3、SCP/28//4、SCP/29/4 和 SCP/30/4。

于涉及使用人工智能的发明，应如何分析创造性步骤以及界定本领域技术人员这一概念。[84]尽管可用的数据有限，但在第三方提出的有关人工智能相关申请/专利的异议中，很多是以缺乏创造性（显而易见性）为理由提出的。[85]

147. 当新技术出现时，创造性评价经常面临特别的挑战。这是因为现有技术参考文献很少，假定的本领域技术人员的确切范围以及该领域的公知常识还未充分确定。缺少判例法和正式指南使主管局难以对创造性作出一致的评价。但是，随着技术趋于成熟，在很多技术领域中已逐渐出现了普遍解释和标准做法。

148. 由于创造性评价是由本领域技术人员作出，因此确定这一假定人员所掌握的知识技能水平就成为创造性评价的基石之一。[86]在每个具体个案中，这一知识技能的确切水平都要得到界定。它也随着技术发展而发生变化。一般来说，假定的本领域技术人员的能力和知识可酌情对应从事各个相关领域技术人员团队的水平。[87]因此，预期的情况是一个人工智能工具在相关领域使用得越多，这种使用的创新性就越低，因为本领域技术人员，即能够使用该人工智能工具的跨学科团队，会在其研究中转而使用这种工具。类似的考量也适用于“公知常识”的概念。[88]

149. 在中国，对涉及算法特征或商业规则和方法的发明专利申请进行新颖性审查时，应当考虑权利要求中说明的所有特征。这些特征既包括技术特征，也包括算法特征或商业规则和方法特征。[89]此外，如果与计算机系统的内部结构有特定技术关系的算法可提高系统的内部性能，那么在评估创造性时也应考虑该算法。同样，如果一个解决方案可通过技术特征，或通过技术特征、算法特征和商业规则或方法的交互作用改善用户体验，那么在评估创造性时也应考虑这种用户体验的改善。[90]例如，如果权利要求中的算法应用于特定的技术领域，能够解决特定的技术问题，则可以认为算法特征与技术特征在功能上相互支持，具有交互关系，并且算法特征成为所采用的技术手段的一部分。[91]

150. 在法国，人工智能领域的专利申请与涉及计算机实施发明的专利申请的审查方式相同。对于混合型发明（即同时包括技术特征和非技术特征的权利要求）的创造性评估，仅考虑为发明的技术特征做出贡献的特征。这可能包括那些单独来看并非技术性，但在发明背景下为产生服务于技术目的的技术效果做出贡献的特征；反之，未对发明的技术特征做出贡献的特征无法支持创造性的存在。[92]

151. 日本特许厅发布的专利和实用新型审查手册附件 A 载有一些关于评价人工智能相关发明创造性的示例。[93]欧洲专利局在创造性方面对人工智能相关发明的评估方式与其他计算机实施发明的评估方式类似。在此，要确保在评估创造性时只考虑有助于实现发明技术特性的技术特征。特别是“非技术性”特征，在这里指的是那些本身属于被排除在可专利性之外的特征，只有当这些特征有助于解决技术问题时，才考虑对其进行评估。如果技术特征对于解决技术问题不发挥作用，那

---

84 见智利答复 C.9199 的评论意见。
85 2019 年世界知识产权组织技术趋势——人工智能，第 115 页至第 117 页。
86 见文件 SCP/22/3。
87 文件 SCP/22/3 第 34 段和第 35 段。
88 见文件 SCP/28/4。
89 中国国家知识产权局审查指南第二部分第 9 章第 6.1.3 条。
90 专利审查指南（2023 年）修改解读（四）——涉及计算机程序的发明专利申请审查。
91 日本特许厅-中国国家知识产权局人工智能相关发明比较研究。
92 见法国答复 C.9304 的评论意见。
93 专利和实用新型审查手册附件 A，示例 31 至 40，日本特许厅。

么在评估创造性时也可能不予考虑。欧专局上诉委员会曾多次做出裁决，涉及与人工智能和机器学习相关发明的创造性评估。[94]

152. 关于人工智能机器“创造”的发明，存在关于人工智能机器大量创造“新发明”的关切，有人担心这会造成所有发明都是由机器创造并获得专利授权的情况。在一定程度上反映了上述关切的一些项目通过公开人工智能机器的输出来产生使用人工智能技术的“现有技术”，使他人无法为上述任何输出获得专利权。[95]

153. 在菲律宾，尽管并未禁止人工智能辅助发明或人工智能生成的外观设计，但人工智能辅助的申请应格外仔细或审慎审查，包括审查人工智能生成输出所依据的材料是否可被视为现有技术。[96]

154. 关于新发明，充分公开要求和工业实用性（实用性）要求可避免对这样的发明授予专利，例如只是已知化学元素的组合，并未说明如何生成这种合成物及其如何使用。同样地，在已公布的引文中所述的信息只能被视为公众可用的信息，因此是适格的现有技术引文，条件是对这些信息进行了充分细致的说明，使本领域技术人员能够实现这些教导。例如仅仅以化学式形式公开的化学结构很有可能不被视为适格的否定对应化学合成物新颖性/创造性的现有技术引文。

155. 创造性（非显而易见性）要求的出发点是不应对可作为公知常识的明显结果推断出的发明给予专利保护，因为它对于社会的贡献微乎其微。[97]这样的政策目标可指导在包括人工智能相关发明在内的个案中对创造性进行判定。

D. 评估可专利性背景下的人工智能生成的现有技术和 PHOSITA

156. 人工智能系统能够持续生成技术公开内容，其规模和速度可能超出已有的现有技术检索、分类和提取系统的能力。[98]此类公开内容可包括与结构生物学、药物发现、分子生物学和生物技术潜在相关的蛋白质结构预测等。[99]由此产生的海量信息可能使申请人、审查员和第三方更难识别技术上相关、可公开获取且足够可靠的公开内容。下文讨论的材料并未表明各司法管辖区已通过专门针对人工智能生成现有技术的立法或行政改革。

157. 不过，欧洲专利局（欧专局）已澄清，人工智能生成的材料可构成现有技术的一部分，如果该材料在相关日期之前已向公众提供，且包含使本领域技术人员能够付诸实践的技术教学。[100]根据这一方式，公开内容的来源、地理位置或形式本身并非决定性。主要考虑因素是向公众提供的技术信息、相关日期以及公开内容的可实施特征。现有的现有技术框架因此可适用于人工智能生成的材料。尽管如此，这些框架在多大程度上能够有效应对人工智能生成公开内容的规模、质量和可靠性仍是悬而未决的实际问题。

---

94 《人工智能和机器学习判例法》：https://www.epo.org/en/legal/case-law/2022/clr_i_d_9_2_11_e.html。

95 所有现有技术项目（https://allpriorart.com/about/）。

96 见菲律宾答复 C.9304 的评论意见。

97 文件 SCP/22/3 第 3 段。

98 知识产权专家组，“应对人工智能生成的现有技术的挑战”，可访问：https://www.ipeg.com/navigating-the-challenges-of-ai-generated-prior-art/

99 Zhidian Zhang、Chenxi Ou、Yehlin Cho、Yo Akiyama、Sergey Ovchinnikov，“用于蛋白质折叠与设计的人工智能方法”，《结构生物学当前观点》期刊，第 93 期，2025 年。

100 欧专局，“数字发明的可专利性”，可访问：https://www.epo.org/en/news-events/in-focus/digital-innovations/patentability-digital-inventions。

158. 其他专利局，包括美国专利商标局（美国专商局），已就人工智能生成的公开内容对现有技术和可专利性判定的可能影响发起磋商。[101]这些讨论确定了若干问题，可能在其他司法管辖区也相关：

- 适用的专利法是否预设现有技术公开内容由人类创作或公布，以及人类贡献的缺失或程度是否影响人工智能生成的公开内容作为现有技术的可用性；
- 哪些形式的人工智能生成公开内容可能与可专利性评估相关，此类公开内容目前如何向公众提供，以及确保公众可获取的额外机制是否合适；
- 提交已知由人工智能生成的出版物或其他证据的一方是否应将此事实通知专利局，以及该方是否应负有确定该公开内容如何生成的义务；和
- 人工智能生成的公开内容是否应与人类生成的公开内容区别对待，特别是考虑到可能的不准确或幻觉，及其对可操作性、可实施性、证据可靠性和公开可获取性的影响。

159. 人工智能系统生成和传播技术信息的规模与速度可对专利制度产生重大实际影响。大量自动生成的材料增加了申请人评估相关现有技术以及审查员检索、分类和评估潜在相关专利和非专利文献所需的时间和资源。这导致在许多情况下相关现有技术淹没在大量无关的人工智能生成现有技术中的风险[102]。因此，一种观点认为，此类人工智能生成的公开内容作为面向工程师和科学家的技术文献，可能价值有限且缺乏证据分量，可能无法反映其实际工作；而且，此类人工智能生成现有技术的规模可能试图排除未来发明，从而给专利制度带来负担。[103]

160. 重复、推测性或技术上不可靠材料的累积也使得相关公开内容更难被找到并在特定情况下对可专利性构成实际障碍，尽管该材料已正式公开提供。不过，在缺乏针对人工智能的专门规定情况下，针对基于互联网公开的一般原则似乎适用。[104]相应地，在线材料可被视为可公开获取，如果公众在相关日期之前能够了解其内容，且不受保密义务或其他使用或传播限制。[105]具体的人工智能生成公开内容是否满足该标准仍需根据具体事实判断，包括其提供的日期、可检索性以及相关公众的可获取性。[106]

161. 一个核心问题是，人工智能生成的公开内容是否构成可实施的现有技术。[107]就新颖性而言，现有技术文献通常仅在其中包含充分信息，结合相关日期的普通常识足以使本领域技术人员能够实施所公开的技术教导的程度时，影响要求保护的发明的新颖性。因此，如果公开内容未提供获得或实施的可行方式，仅提及概念、结果、化合物或结构特征可能不够。不一定要求实际实施或事先实证测试。如果公开的客体能够由本领域技术人员付诸实践，则理论性或预测性人工智能生成公开内容仍可能与新颖性相关。在评估创造性时，调查有所不同，取决于可适用的国家或地区

---

101 关于人工智能激增对现有技术、本领域普通技术人员知识以及基于上述做出可专利性判断的影响征求意见 89 FR 34217，2024 年 4 月 30 日，案号 PTO-P-2023-0044。

102 BIO 针对美国专商局征求意见的评论（89 FR 34217，2024 年 4 月 30 日，案号 PTO-P-2023-0044）。

103 同上。

104 Tarrere 等（2026 年），“在欧专局质疑人工智能生成现有技术的思路”，可访问：https://www.santarelli.com/en/document-genere-par-ia-contester-oeb/

105 同上。

106 同上。

107 Holman 等（2024 年），“人工智能生成的现有技术对专利法的五大潜在影响”，可访问：https://www.sternekessler.com/news-insights/insights/top-5-potential-implications-of-ai-generated-prior-art-on-patent-law/

方法。相关问题包括本领域技术人员在寻求达成要求保护的发明时，是否会将公开的教导视为可信且在技术上有意义的起点，或者是否会将其纳入考虑。

162. 在确定可实施性和证据可靠性方面，人工智能生成的公开内容尤其带来困难。[108]人工智能生成的输出不会透露为其生成做出贡献的模型参数、算法、训练数据、提示或其他条件，这使得确定公开的结果是否可复制或查明任何未能复制的原因非常困难。此外，人工智能输出可能包含幻觉、事实不准确、内部不一致的陈述或不精确的技术描述。[109]这些特征并不一定意味着人工智能生成的公开内容不具备可实施性；每项公开都需要基于其内容，并结合本领域技术人员的普通常识评估。不过，当大量推测性、预言性、未经核实或技术上前后矛盾的材料自动放入公有领域时，这些特征可能尤为重要。这引发了一个尚未解决的政策问题，即关于公开可获取性、证据分量和可实施的公开内容的现有规定是否提供了足够的手段，区分技术上有意义的现有技术与仅描述缺乏支撑或无法操作的可能性的材料，或者进一步明确这些规定是否会有帮助。

163. 现有方式表明，公开内容的来源——无论是人类还是人工智能生成，通常并非其作为现有技术的地位或相关性的唯一决定因素。根据可适用的法律，相关考虑因素包括能否确定公开日期、信息在相关日期前是否已可公开获取，以及技术教导是否具有充分的可实施性。提供技术上可复制的教导的人工智能生成公开内容因此可能影响新颖性，或与创造性评估相关，而单纯的概念、缺乏支持的预测或技术上不可操作的公开则可能不足以满足上述目的。尽管特定法律框架要求在其他情况下标注人工智能生成内容，但审查的材料并未表明此类标注要求已普遍纳入涉及现有技术识别或提交的专利法规定中。[110]尽管如此，提交人工智能生成参考资料的当事人和从业人员仍须遵守适用的合理调查、坦诚、善意或职业责任义务。

164. 人工智能工具的日益普及和成熟引发了关于本领域技术人员（在一些司法管辖区亦称为本领域普通技术人员（PHOSITA））的法律建构问题。这一假设性标准与专利审查的多个方面相关，包括出于新颖性目的对权利要求和现有技术的解释、对创造性或非显而易见性的评估，以及确定发明是否已充分公开，使本领域技术人员能够实施。[111]特别是，人工智能在研究和开发中的使用增加，可能影响哪些活动会被视为常规，而哪些仍继续需要创造性技能。因此，核心问题并非人工智能是否总体上使发明更具有或更缺乏创造性，而是具体的人工智能辅助活动在可适用的日期，是否已成为相关技术领域内技术人员通常可获得并常规执行的活动[112] 。

165. 现有的国家和地区做法并未表明，专门针对人工智能重新界定 PHOSITA 或归属于该人员的普通常识（CGK）已得到普遍采纳。[113]当前的方式很大程度上保留了以人为中心的法律建构，将人工智能视为本领域技术人员可用的工具，而非该假想人物的构成要素。因此，使用具体人工智能工具是否构成技术人员普通能力的一部分，将取决于该工具在适用日期于相关领域内的可获取性及惯常使用等其他考虑因素。对这些能力的准确界定具有横向影响，可能影响创造性、权利要求的

---

108 Tarrere 等（2026 年），“在欧专局质疑人工智能生成的现有技术的思路”，可访问：https://www.santarelli.com/en/document-genere-par-ia-contester-oeb/

109 Mehta（2024 年），“现有技术问题：法律顾问审视人工智能生成的公开内容”，可访问：https://www.managingip.com/article/2d6pfurgj9will0moq2o0/patents/prior-art-problems-counsel-mull-ai-generated-disclosures

110 中国的人工智能监管：2025 年标注规定的临时措施，可访问：https://casrai.org/guides/china-ai-regulation-explained

111 Roberto Dini、Mario Franzosi，“所谓的‘本领域技术人员’：是驴还是天才？”，《知识产权法与实务杂志》，第 19 卷第 5 期，2024 年 5 月，第 387－389 页；和 Roberto Dini，“人工智能创新与‘本领域技术人员’”，《知识产权法与实务杂志》，第 20 卷第 10 期，2025 年 10 月，第 643－644 页。

112 产权组织人工智能与知识产权政策工具包，第 15 页。

113 同上。

解释、占先或新颖性的评估，以及在可适用的法律相关的情况下，书面说明、可实施性和公开充分性。

166. 在此背景下，可以将假设的本领域技术人员的法律身份或性质，与归属其的知识、工具和常规能力区分开来。虽然目前没有司法管辖区针对人工智能带来的挑战发布 PHOSITA 指南，但美国专商局已启动磋商，这将在下文讨论。[114]此类问题在概念上与总体专利制度相关，在其他司法管辖区可能带来相似的挑战。这些问题包括

- 本领域技术人员是否应继续被理解为以人类为中心的建构，或当技术性质有此要求时，被理解为跨学科团队；以及是否可假设这两种建构使用人工智能工具，而不将人工智能本身视为法人；
- 人工智能工具的广泛提供和日常使用如何影响归属于技术人员的普通常识、权利要求术语的解释以及为创造性目的对类似现有技术的识别；
- 特定人工智能系统的概率性和非确定性运行如何影响对可预测性和成功合理预期的评估；
- 人工智能辅助检索能力在出于新颖性和创造性目的确定现有技术时是否以及在何种程度上应纳入考量；
- 在有效申请日期，人工智能系统的版本、功能、训练数据及可用信息应如何在时间上与技术人员的特征一致；和
- 对人工智能辅助实验和优化工具的获取如何影响可实施性和公开充分性，包括全面实施要求保护的发明所需的实验，对于本领域技术人员是合理还是过度。

167. 在欧洲专利组织的框架内，尽管对 PHOSITA 的方式仍以人为中心，但材料显示类似的关切已欧洲专利组织缔约国之间的讨论中得到承认，而关于机器能否被认定为技术人员的政策问题仍悬而未决。认识到，人工智能的使用可能会增加技术人员的知识，但这难以确立。此外，还认识到，技术人员必须是能够运用人工智能的跨学科团队。[115]

168. 在特定的人工智能相关发明中，本领域技术人员可适当由将人工智能技术专门知识与相关应用领域专门知识结合的团队代表。例如，评估用于安全和监控的基于人工智能的图像识别系统需要同时具备机器学习技术和监控技术的知识。这并不一定构成对人工智能增强法人的认可；而是反映了已确立的可能性，即根据技术问题的不同，技术人员可以由多学科研究或生产团队代表。[116]随着标准的人工智能工具在特定领域得到常规使用，一定程度的人工智能素养或使用此类工具的能力，正日益成为归属该人员的普通能力或普通常识的一部分。不过，任何此类结论均应基于相关工具在适用日期于该领域内的实际可获取性和采用的证据，而非基于其后续发展或在其他技术领域的使用。[117]

---

[114] 关于人工智能激增对现有技术、本领域普通技术人员知识以及基于上述做出可专利性判断的影响征求意见 89 FR 34217，2024 年 4 月 30 日，案号 PTO-P-2023-0044。

[115] 欧洲专利组织，“涉及人工智能（AI）的发明专利的法律方面：欧洲专利公约缔约国的反馈摘要”，可访问：https://link.epo.org/web/AI_inventorship_summary_of_answers_en.pdf

[116] Steadman 等（2026 年），“人工智能增强 POSITA：人工智能如何重塑美国专利法中的显而易见性审查”，可访问：https://www.dlapiper.com/en/insights/publications/intellectual-property-news/2026/ai-augmented-posita-how-artificial-intelligence-is-reshaping-obviousness-inquiry

[117] Block 和 Ratz（2021 年），“扼杀在现有技术中？人工智能如何挑战专利法中技术人员的虚构概念”，可访问：https://lesi.org/article-of-the-month/killed-in-the-arthow-artificial-intelligence-challenges-the-fictional-concept-of-the-skilled-person-in-patent-law/

169. 人工智能模型和工具的快速持续发展带来了另一项难题。[118]若创造性论点假设本领域技术人员会使用人工智能来识别、选择或结合来自两个或多个现有技术文献的教导，则有必要确定该人员在优先权日或有效申请日期（如适用）实际具备哪些能力。依赖人工智能工具晚些时候或经实质改进版本的能力，可能对分析引入事后视角，特别是当后续版本在检索、推理、预测或优化能力上存在实质性差异时。因此，涉及发布日期、模型或版本信息、访问条件、训练数据截止点以及同期使用的证据可能辅助评估。在合适的情况下，数字记录、日志或带时间戳的信息或许也相关。[119]不过，它们的潜在证据价值以及可能要求它们的具体情形仍需进一步考虑，不应被理解为对申请人确立了记录保存的一般性义务。[120]

170. 在现行做法下，PHOSITA 通常仍是以人为中心的法律建构，人工智能被视为该人员或跨学科团队（酌情）可用的工具。因此，仅使用人工智能或软件来实施此前由硬件执行的功能、将常见的软件程序转移到另一个领域以实现相同功能，或者计算机实施已知由人类开展的商业方法，若仅需常规系统分析或调整，均属于普通技术。[121]例如，仅仅是将人工智能应用于一个技术领域，通过示例提供指导的专利局通常认为，以软件的方式实施由电路等硬件开展的功能的尝试，属于本领域技术人员的正常工艺能力。同样，将一个特定领域的通用软件程序或手段应用于另一个领域，功能或操作相同的，视为属于技术人员的常规能力范围。如果引用的发明公开了人类开展的商业服务或方法，而非其系统化，则计算机实施该发明仍属于技术人员的普通创造能力，前提是能够通过常规系统分析和设计实现。[122]不过，这一评估取决于在有效申请日期相关领域中可合理获得的能力和工具，因为人工智能的日益日常提供会影响归属 PHOSITA 的基础知识和能力，但并不一定改变可适用的法律标准。[123]

171. 关于训练数据的修改，评估可区分常规选择与产生意料外技术效果的配置。一般预期与理想输出相关联的变量的选择，包括为提高机器学习系统的可靠性或准确性，将构成普通常识的一部分，因此无法支持创造性。相反，若输入数据基于先前未知的相关性进行组合，且特定的训练参数产生出乎意料的技术效果，则可能存在创造性。[124]欧专局已允许数据转换相关的权利要求（例如，在与涉及将许多图像处理成一张的现有技术对比时，将一张图像处理成多张的发明）具有创造性。[125]在从 PHOSITA 的角度评估创造性时，可能还需要考虑人工智能系统的概率性和非确定性特征，特别是在确定技术人员是否对实现所主张的结果具有合理预期时。[126]

172. 同样，尽管通过训练数据的预处理提高神经网络的准确性属于常规目标，但特定的预处理修改不应自动被视为显而易见。若修改既未在现有技术中公开，也不属于普通技术常识的一部分，

---

[118] 同上。
[119] 同上。
[120] 同上。
[121] 日本特许厅《专利和实用新型审查手册》，附件 B 第 1 章第 2.2.3.3 节，见示例 1、2、4、5 和示例 5]，以及《专利和实用新型审查手册》，附件 A，见案例 33、34、37、38 和 40。另见大韩民国知识产权部，“按技术领域分列的审查做法指南”[第 6 章，案例 8.11.14]、[案例 34（权利要求 1）]、[案例 37]、[案例 38]、[案例 40]。
[122] 同上。
[123] Stephen How、A Hou、Stephen M.，“国际专利局关于人工智能发明的指南”（2024 年），可访问：https://www.solveintelligence.com/blog/post/international-patent-office-guidance-on-artificial-intelligence-inventions。
[124] 日本特许厅《专利和实用新型审查手册》，附件 A[案例 34（权利要求 2）]和[案例 35]，以及韩国知识产权部“按技术领域分列的审查做法指南”[第 1 章案例 4][和[第 6 章案例 10.12.13]。
[125] T 1286/09（图像分类器/INTELLECTUAL VENTURES），2015 年 6 月 11 日（欧专局）。
[126] Stephen How、A Hou、Stephen M.，“国际专利局关于人工智能发明的指南”（2024 年），可访问：https://www.solveintelligence.com/blog/post/international-patent-office-guidance-on-artificial-intelligence-inventions。

且产生了超出常规优化的显著、意想不到的技术效果，则将产生创造性。[127]因此，评估将依次考虑修改的性质、相关现有技术、技术人员的普通常识、特定训练配置是否显而易见以及获得的技术效果，同时考虑到在相关日期 PHOSITA 能够合理获得的工具和实验能力。

173. 最后，若深度学习训练和使用经训练的神经网络的估算已十分知名，且现有技术解决了相同问题并执行相同功能，将预期技术人员会应用该教导。在这种情况下，贡献仅在于使用已知的深度学习方法获得可预见的估算改进的要求保护的发明将不具备创造性。[128]

E. 充分公开和权利要求[129]

174. 与创造性评价相同，新技术对于发明的公开提出了特别的挑战，要求清楚完整地公开发明，以及清晰扼要地撰写权利要求，以充分涵盖法律保护的范围。缺少判例法和正式指南使知识产权局和专利制度的用户同样难以对是否符合公开要求进行评价。

175. 关于要求保护发明的说明书，一般来说，国家/地区专利法要求专利申请人应以充分清晰完整的方式对发明进行公开，使本领域技术人员能够实现要求保护的发明（充分公开要求）。[130]专利制度正是通过这一要求为传播信息以及获取专利申请和专利所包含的技术知识提供便利。这使技术知识在公众中的积累得以扩大，并增进了社会总体效益，例如促进技术转让和避免重复性的研发工作。

176. 在人工智能技术方面，一个可能的问题是人工智能算法、训练模型、神经网络架构、学习过程、训练数据、硬件组件等信息在专利申请中应以怎样的程度公开才能符合充分公开条件。[131]因此，评估人工智能发明是否充分公开是一个新的挑战，许多国家尚未就此确立判例法。[132]挑战之一可能是由于就目前的深度学习技术而言，人们难以识别深度学习神经网络中所执行的每个流程步骤，以及明确解释神经网络如何得到最终结果。如果一个系统有数千万个用于分类的权重，则该系统会过于复杂，以至于无法以人可以理解的方式对其进行表述。在某些情况中，在没有实际试验数据的情况下对人工智能输出进行合理化解释（即以可信的方式进行论证）可能更为困难。

177. 与此同时，要求保护的发明在专利申请说明书部分的公开程度显然取决于在申请的权利要求部分中要求保护的内容。例如一项发明是关于通过利用特定数据集训练深度学习算法来应用人工智能技术解决一个问题，如果要求保护的发明有更广泛的应用，则可能要求在说明书中公开本领域技术人员实现要求保护的发明广泛应用所必需的所有数据集类型，而不是只公开一种数据集类型。

178. 在这方面，本领域技术人员的概念对于评价充分公开同样重要。例如，如果一项人工智能技术被应用于某一领域的发明（如图像识别神经网络应用于安防和监控领域的发明），则人工智能技术领域和监控领域技术人员团队可作为假定的本领域技术人员来对这项发明进行评价。

---

127 日本特许厅《专利和实用新型审查手册》，附件 A[案例 36]和“按技术领域分列的审查做法指南”[第 6 章案例 16]

128 日本特许厅《专利和实用新型审查手册》，附件 A[案例 32]和大韩民国知识产权部“按技术领域分列的审查做法指南”[第 1 章案例 2.3.5]

129 另见文件 SCP/34/5（关于充分公开的进一步研究（第一部分）），四、人工智能相关发明。

130 见文件 SCP/22/4。另见“国家/地区专利法的若干方面——充分公开”：https://www.wipo.int/scp/en/annex_ii.html。

131 见智利和德国答复 C.9199 的评论意见。

132 例如，见德国答复 C.9199 的评论意见。

179. 另一个问题可能是由于深度学习技术的非确定性：它们需要进行一些随机的初始化。因此，相同的训练数据和相同的神经网络架构也可能使机器学习有略微不同的绩效。使用相同训练数据和相同神经网络架构对模型进行训练会产生两种略有区别的训练行为。与生物材料不可避免地存在生物差异性的情况相同，可以基于专利申请中的公开考虑要求保护的发明所谓的可再现性和可行性。

180. 关于训练数据，通过一项特定人工智能技术解决一个问题可能需要一个特定数据集。训练数据集在深度机器学习的绩效方面所发挥的重要作用可能会让人们提出这样的问题，即此类数据集在专利申请中的公开程度应为多少，以及是否应提供此类数据集以供第三方对要求保护的发明进行验证（即要求保护的发明是否确实起作用）。

181. 关于权利要求，很多国内法规定权利要求应清晰扼要。此外，权利要求应得到说明书的支持（支持要求）。[133]一般来说，这项要求的出发点是要求保护的发明不应超出说明书中所公开披露发明的范围。同样地，美国法律所规定的书面说明书要求的关键政策目标[134]是“明确传达申请人发明了要求保护的主题的信息，以及使公众掌握申请人要求保护发明的信息”。[135]因此，这些要求体现了这样的基本原则，即对于申请人截至申请日并未发明的内容以及截至申请日并未通过专利申请中的公开与公众分享的内容不能给予专利保护。由于人工智能相关发明大多是计算机实施发明，在人工智能相关发明的权利要求撰写方法方面，申请人可能面临相同的挑战，即在权利要求中要为其发明确定适当的保护范围。

182. 在中国，2025 年修订的《专利审查指南》明确指出，公开充分性的相关要求在两种涉及人工智能的情境下已满足：“如果涉及人工智能模型的构建或者训练，则一般需要在说明书中清楚记载模型必要的模块、层级或者连接关系，训练必需的具体步骤、参数等；如果涉及在具体领域或者场景中应用人工智能模型或者算法，则一般需要在说明书中清楚记载模型或者算法如何与具体领域或者场景相结合，算法或者模型的输入、输出数据如何设置以表明其内在关联关系等，使得所属技术领域的技术人员按照说明书记载的内容，能够实现该发明的解决方案”。中国进一步指出，经修订的《指南》包括两个案例，说明相关条款的适用标准。[136]

183. 在法国，涉及人工智能领域专利申请公开充分性的指导原则目前正在起草，预计最迟将于 2026 年底前在网上发布。[137]

184. 在印度，2025 年 CRI 指南对人工智能/机器学习/深度学习发明提供了具体的公开指导。强调人工智能/机器学习/深度学习权利要求必须有充分的细节支持，使本领域技术人员能够不经不必要的试验即可复制技术解决方案。取决于所主张的贡献，申请人应公开神经网络架构和结构详细信息、训练数据特征及标注方法、预处理步骤、如损失函数、优化器设置和训练轮次等训练细节、验证指标和实际测试结果、输入-输出关联以及数据管道或算法架构（如相关）。如果创造性取决于训练数据集的特定特质，指南称这些特质应公开，除非无需进行不必要的实验即可识别，尽管具体的数据集本身通常无需公开。[138]

---

133 见文件 SCP/22/4。
134 美国法案第 35 章 112(a)条。见文件 SCP/22/4。
135 同上。
136 见中国答复 C.9304 的评论意见。
137 见法国答复 C.9304 的评论意见。
138 专利、外观设计和商标总署，2025 年《计算机相关发明（CRI）的审查指南》，第 5.2.2 节

185. 关于人工智能相关发明对于公开要求的适用，日本特许厅发布的专利和实用新型审查手册附件 A 载有若干具体示例。[139]在大韩民国，说明书应详尽解释人工智能相关发明的具体手段、解决的技术问题和提供的解决方案，确保技术人员能够轻松理解和再现该发明。如果一项发明没有具体规定经过训练的模型的输入数据和输出数据之间的相关性作为实现人工智能相关发明的特定手段，则该发明被视为公开不充分。经过训练的模型的输入数据和输出数据之间的相关性是指：(i)具体说明训练数据；(ii)要求保护的发明各特征之间存在解决技术问题的相关性；(iii)具体说明一种通过使用训练数据进行训练的学习模型或一种训练方法；(iv)通过这种训练数据和训练方法生成用于解决要求保护的发明的技术问题的训练模型。然而，如果本领域技术人员，在将申请提交时的公知常识考虑在内的情况下，能够根据发明说明书中所述的模型推测或理解这种相关性，则视为满足了能够实现的要求。此外，仅使用功能框图或一般流程图而不提供具体实施细节通常是不够的。申请必须说明硬件或软件如何以本领域技术人员可以理解和再现的方式实现发明的功能。这种详细的公开可确保发明能被有效地再现和利用，从而满足能够实现要求，并支持创新的实际应用性。[140]

186. 2024 年 3 月，欧洲专利局更新了审查指南，对人工智能相关发明的充分公开要求作出了实质性修改。[141]更新后的指南中的一项重要修订是，要满足充分公开要求，可能必须在专利申请中提供关于要求保护的人工智能模型的训练数据的详细信息。其目的是确保充分公开，并作为技术效果的证据。具体来说，更新后的指南指出，下列情况下专利公开不充分，即如果“数学方法和训练数据集的公开不够详尽，无法在要求保护的整个范围内再现技术效果。缺乏详细信息的公开可能使得专利公开更像是对研究计划进行征集”。[142]新指南还指出，“机器学习算法实现的技术效果可能是显而易见的，或是通过解释、数学证明、实验数据或类似方法所确立[……]如果技术效果取决于所使用的训练数据集的特定特征，则必须公开重现技术效果所需的这些特征，除非技术人员能够利用公知常识确定这些特征，而不会造成不必要的负担。不过，一般来说，没有必要公开具体的训练数据集本身”。[143]尽管该指南目前规定，一般不要求完全公开基础训练数据集，但在哪些情况下必须公开数据集尚未完全确定。

187. 欧洲专利局做出了两项与充分公开和人工智能有关的决定。在 T 0161/18 案中，欧洲专利局上诉委员会认定，该案中涉及人工神经网络的专利申请不符合充分公开的要求。上诉委员会指出，该申请没有公开哪些输入数据适合训练人工神经网络，也没有公开至少一个适合解决该技术问题的数据集。[144]同样，在 T 1191/19 案中，一件专利申请也被认定不符合充分公开的要求。在得出这一结论时，委员会注意到该申请没有提供任何训练数据的示例，甚至没有提供使该发明能够提供有意义的预测所需的训练数据量。[145]这些案例强调，申请人必须确保其公开的信息能够实现，可以包括详细的训练数据和其他相关信息，以符合欧洲专利局的要求。

---

139 专利和实用新型审查手册附件 A，示例 46 至 51，日本特许厅。
140 《人工智能领域审查指南》，韩国知识产权局。
141 《欧洲专利局（欧专局）审查指南》，G 部分，第二章，3.3.1。
142 《欧洲专利局（欧专局）审查指南》，F 部分，第三章，公开不充分。
143 《欧洲专利局（欧专局）审查指南》，G 部分，第二章，3.3.1。
144 欧洲专利局上诉委员会第 T0161/18 号决定。
145 欧洲专利局上诉委员会第 T1191/19 号决定。

188. 菲律宾知识产权局（IPOPHL）也指出，不可靠的来源或虚构来源是人工智能辅助发明中的常见问题。一旦发现此类来源，审查员虽无需核查说明书中引用的每一项参考资料，但应在审查意见书中提出该问题，以便删除或更正此类资源，以维护菲律宾授予的发明的真实性和质量。[146]

189. 对不同司法管辖区的上述调查显示，作为一个新的技术领域，对人工智能相关发明适用充分公开要求仍在演变之中，对此尚未有明确的规定。一些司法管辖区正在制定报告和规则，以应对这些特殊的挑战。[147]该领域立法、指南和法院裁决的不断发展将有助于确立更为准确的标准和做法。从一般意义上讲，许多人工智能算法固有的黑箱现象可能不会对发明的充分公开造成问题，只要提供足够的详细信息说明使用哪种人工智能算法以及如何训练即可。[148]

190. 值得注意的是，在许多人工智能模型和基于人工智能的发明中，要求保护的发明本质在于一种技术关系，该关系将特定类别的输入数据（例如医学影像、传感器读数或文本数据）与特定输出或结果（例如诊断、预测或分类）关联起来。[149]如果这种相关性无法通过普通常识理解，说明书应提供具体的依据，使技术人员能够理解。[150]换言之，说明书必须提供充分依据，使读者能够得出结论认为所主张的相关性确实存在且在技术上具有充分依据，而非仅仅是断言。[151]

191. 仅陈述相关性存在，而不提供任何支持性理由，不足以满足这一要求。说明书需要证实要求保护的相关性，并且这可以通过解释该相关性为何成立的理论依据，或者通过提供实证数据、统计结果、实验结果或说明书中包含的其他形式技术依据实现。这一要求涉及公开本身的充分性和质量，是独立于权利要求范围相关问题的考虑因素。[152]

192. 一个密切相关的问题涉及说明书是否包括从真正构建并经过测试的人工智能模型中获得的实证性能数据，例如准确率、精确率和召回数字，或与公认基准的比较，以证明所主张的相关性。此类数据的存在有助于区分仅在概念上描述的人工智能系统与已证明如权利要求所述实际表现的人工智能系统。这进而有助于防止底层模型从未实际实施或测试过的投机性或纯粹预言性的人工智能权利要求。[153]

193. 输入数据的预处理，可能涉及数据清洗、归一化、特征提取、转换和增强，构成许多基于人工智能的发明中具有技术重要性的部分。如果这些预处理步骤仅以模糊或笼统的措辞描述，说明书有不满足可实施性要求的风险，因为本领域技术人员无法仅基于此复制发明。同样，说明书需要以足够具体的措辞描述使用的学习模型，包括神经网络的类型、层数、激活函数和相关超参数等细节，以及训练方法，包括学习算法、损失函数和采用的优化策略。[154]如果笼统提及无法使技术人员复制专门针对主张的技术问题的模型，则可能不充分。[155]

194. 多个司法管辖区已明确指出，根据所涉发明的性质，人工智能发明的公开要求应明确说明基本模型结构、涉及的训练步骤、相关参数以及关联输入数据与输出数据的关系，以便本领域技

---

146 见菲律宾答复 C.9304 的评论意见。
147 见巴西答复 C.9199 的评论意见。
148 见德国和葡萄牙答复 C.9199 的评论意见。
149 见 SCP/34/5，第 27 页、第 32－33 页。
150 同上。
151 日本特许厅《专利和实用新型审查手册》，附件 A[案例 46]、[案例 47]和[案例 48]，以及 MOIP，大韩民国“按技术领域分列的审查做法指南”[第 1 章案例 1]、[第 6 章案例 2.6.7]。
152 日本特许厅《专利和实用新型审查手册》，附件 A[案例 49]。
153 日本特许厅《专利和实用新型审查手册》，附件 A[案例 50]。
154 MOIP，大韩民国“按技术领域分列的审查做法指南”[第 1 章案例 1]、[第 6 章案例 4]。
155 SCP34/5，第 27 页。

术人员能够实施该技术解决方案，无需开展不必要的实验。[156]仅靠功能图或一般流程图被认为是不够的。至少有一个司法管辖区的判例法已确认，如果一项涉及人工神经网络的专利申请未公开哪些输入数据适用于训练该网络，或者未确定至少一个能够解决该技术问题的数据集，则该申请不符合充分性要求。同样，未提供训练数据示例，也未说明实现有意义预测所需的最小训练数据量的申请，被认定不符合可实施性要求。

195. 训练数据带来了一系列独特的公开相关挑战。当发明的技术性能取决于特定数据集时，自然会产生以下问题：在申请中必须公开多少关于该数据集的信息，以及是否应允许第三方访问该数据集，以便验证要求保护的发明是否如描述的那样运作。要求全面公开训练数据存在重大实际困难，因为此类数据通常属于专有信息且作为商业机密对待。另一个实际问题涉及隐私和数据保护法，特别是当相关数据虽可根据有限许可向发明人或申请人提供用于训练目的，但不得用于公开与基于特定数据集训练的要求保护的人工智能模型相关信息的情况。在现行专利局指南中，这些问题在很大程度上仍未得到解决。[157]

196. 尽管目前并不要求全面公开训练数据，但如果申请甚至未公开一个合适的训练数据集，或未说明发明按权利要求运行所需的最低训练数据量，仍可能被认定为公开不充分。当发明产生的技术效果取决于所用数据集的特定特征时，这些特征必须公开，除非技术人员能够不费过多努力独立确定这些特征。话虽如此，通常认为无需公开数据集本身的全部内容。

197. 已确定用于验证要求保护的人工智能性能的一个可能机制涉及通过向指定机构交存或在专利授予后由第三方验证，实现训练数据的部分或全部公开。但是，这种方法引发了关于不同司法管辖区对已提交数据的跨境访问和验证的进一步问题。虽然与微生物保藏相关的类似问题已通过现有条约框架下的国际合作解决，但涉及训练数据的情况有所不同，因为所涉客体是无形的，原则上能够普遍获取，这意味着实体交存并非严格必要。因此，在此背景下的核心挑战不在于实体可获取性，而在于为评估要求保护的发明可实施性之目的验证训练数据。

198. 虽然并非所有司法管辖区均已解决申请人提交的虚构现有技术及其对充分公开的影响问题，但至少一些专利局对存在不可靠或捏造的来源提出关切，包括在说明书中出现的所谓人工智能生成的虚构内容，已是借助人工智能工具编拟的专利申请中反复出现的问题。[158]若出现此类问题，预期审查员通过适当信函在审查过程中提出，并要求申请人更正或删除有问题的内容，以便维护专利的整体质量及其所含证据的真实性。[159]例如在美国，美国专商局已明确指出，现有的职业义务适用于人工智能辅助提交的申请，尽管不强制施加公开具体人工智能工具使用的普遍义务，除非专门要求。[160]

199. 鉴于公开模型权重或训练数据集要求的模棱两可，且司法管辖区仍在就申请人应如何将其纳入提供指导，学术评论家建议引入与针对微生物保藏的《布达佩斯条约》框架相似的国际交存制度，或科学期刊采用的当地寄存系统。[161]

---

156 SCP/34/5，第 34 页；和 Mateo Aboy 等，“人工智能发明的公开充分性”，《知识产权法与实务杂志》第 19 卷第 11 期，2024 年 11 月，第 836 页和第 838 页。

157 产权组织人工智能与知识产权政策工具包，第 15 页。

158 Hughes（2025 年），“专利行业中人工智能的使用：虚构的阴影”，可访问：https://ipkitten.blogspot.com/2025/10/use-of-ai-in-patent-industry-spectre-of.html

159 见菲律宾答复 C.9304 的评论意见。

160 《关于在美国专利商标局实际使用基于人工智能工具的指南》[案卷号：PTO-P-2024-0013]。

161 Mateo Aboy 等，“人工智能发明的公开充分性”，《知识产权法与实务杂志》第 19 卷第 11 期，2024 年 11 月，第 839 页。

200. 综上所述，适用于人工智能发明的充分性标准显然仍处于持续发展中。话虽如此，许多人工智能系统所具有的黑箱性质本身并不必然导致申请无法满足充分性要求，前提是申请书充分解释了底层算法和相关的训练过程。预计未来无论是通过立法、审查指南还是判例法，在这一领域中适用的标准将逐步完善。[162]

F. 工业实用性

201. 关于要求保护发明的可再现性和可行性，在一些国家，符合工业实用性要求可能也意味着要求保护的发明在必要时必须能够再现且再现结果具有相同的特性。[163]

202. 工业实用性的一个关键方面是要求保护的发明的可再现性和一致性。对于人工智能系统，特别是那些应用深度学习和神经网络的系统来说，由于其固有的复杂性以及决策过程不透明性，要证明绩效的一致性可能具有挑战性。

203. 举例来说，用于工业机械预测性维护的人工智能系统只有在能够持续准确预测故障的情况下，才可被视为具有工业实用性。这可能需要详细记录人工智能模型，包括其训练数据、算法和验证结果，以确保它能在工业环境中可靠实施并产生一致的结果。

G. 发明人权利和所有权

204. 随着人工智能日益融入发明过程，一些重要问题开始显现，涉及人工智能创造之发明的所有权，以及在有人工智能协助的情况下自然人列为发明人的标准。[164]SCP 在 2023 年制定了文件 SCP/35/7，以解决在人工智能相关发明的背景下不断演变的发明人权利和所有权问题。该文件深入概述了人类与人工智能在发明过程中的交互关系，并透彻分析了当前围绕人工智能发明人权利的法律框架和政策问题。[165]因此，本节仅概述关于人工智能相关发明的发明人权利/所有权问题。

专利法规定的发明人权利和所有权

205. 《巴黎公约》第 4 条之三规定，发明人有权要求在专利中提及自己是发明人。该条是指发明人享有在巴黎联盟的所有国家在为其发明授予的专利中记录自己是发明人的权利，通常称之为署名权。普遍的理解是发明人可以放弃这项权利，除非国家立法另有规定。由于《巴黎公约》没有给出“发明人”一词的定义，因此发明人的确定以及行使该署名权的程序由各成员国在其适用法中作出规定。[166]如果一名以上的发明人共同创造了一项发明，则他们是共同发明人。

206. 尽管可专利性要求（如可授权主题、新颖性、创造性（显而易见性）、工业实用性（实用性）和公开要求）不涉及发明人权利的问题，但不准确填写发明人姓名可能产生严重的法律后果。发明人指定不准确会导致各种法律后果，不同司法管辖区对这种不准确的问题提供救济措施。

207. 尽管并非所有国内法都规定了“发明人”的定义，但考虑到专利制度的出发点以及署名权是与专利权相关的基本权利之一，可能存在的普遍推定是专利法下的（一个或多个）发明人是指

---

162 SCP/34/5，第 34 页，另见 Mateo Aboy 等，“人工智能发明的公开充分性”，《知识产权法与实务杂志》第 19 卷第 11 期，2024 年 11 月，第 838-840 页。

163 SCP/5 非正式文件（工业实用性/实用性要求在国家和地区法律下的实际应用）。另见欧洲专利局上诉委员会判例法，I.E.2.部分。

164 见智利答复 C.9199 的评论意见。

165 委员会在 SCP 第三十五届会议上商定，秘书处将更新文件 SCP/35/7 第五和第六部分并提交 SCP 第三十七届会议。

166 《保护工业产权巴黎公约申请指南》，G.H.C. Bodenhausen（产权组织第 611 号出版物）。

（一个或多个）个人。[167]如果这是一个正当推定，那么逻辑结论可能就是无论人工智能机器为发明的构想作出了多大程度的贡献，机器不能作为发明人。[168]

208. 如果在发明创造的过程中使用了人工智能系统，只要在这一过程中有一人（或多人）符合适用法下“发明人”的条件——宽泛地说就是为构想要求保护的发明作出了贡献——则该人（或这些人）就是该发明的发明人，发明人可以是人工智能程序员、人工智能开发人员、人工智能用户或其他人员。这里理论上存在的一个问题是如果没有人符合适用法下发明人的条件，谁享有对于专利的权利？

209. 在印度，2025 年 CRI 指南明确区分了人工智能生成的发明与人工智能辅助发明。指南称，人工智能生成的发明，即自主创造或仅在非常有限的人工干预下产生的发明，不具备可专利性，因为根据 1970 年《专利法》第 6 条，人工智能不得被称为主张是真正、第一发明人的“人”。相比之下，人工智能辅助发明根据第 3(k)条并非一概不具有可专利性，前提是符合可专利性标准并且通过实体化应用展现出技术效果。[169]

210. 同时，菲律宾指出，在符合伦理考量的前提下，人工智能可作为辅助发明人、制造人或设计师的工具；当个人对人工智能系统使用提示或指令协助生成发明或外观设计时，该个人仍可被视为发明人和/或申请人，因为根据现行菲律宾法律，发明和外观设计仍与人类创造和智力贡献相关联，而人工智能系统本身不得在申请中列名。[170]

211. 2025 年 11 月，美国专商局发布了《关于人工智能辅助发明的发明人资格修订指南》，完全废除了其 2024 年 2 月关于该主题的指南。新指南强调了长期以来的先例，即决定发明人资格的同一法律标准适用于所有发明，无论发明过程中是否使用人工智能系统，且不为人工智能辅助发明制定或适用新的、单独或经修改的标准。在美国，只有自然人能被适当列为专利申请的发明人：《美国法典》第 35 编第 100 条第(f)项规定，“发明人”指发明或发现发明客体的个人，而《专利法》中的“发明”指发明人的构思。美国专商局推定，申请数据表或宣誓书/声明中列名的发明人为真实的人类发明人，而包括生成式人工智能和其他计算模型在内的人工智能系统是人类发明人使用的工具，既不符合发明人资格，也不能将此类辅助提升至发明人地位。[171]

212. 尽管随着技术的进步，预计人工智能机器将具有更强的认知能力，但技术发展往往是增量式发展。此外，人工智能技术在发明创造过程中根据不同的情况可能发挥不同的作用，即可能发挥从只是作为辅助工具到帮助构想发明概念的手段的各种作用。因此，把“人类创造的发明”和“机器创造的发明”对立起来对于关于发明人权利问题的复杂讨论来说似乎过于简单。

---

167 根据 35 U.S.C. § 100(f)，“发明人”是指“发明或发现了发明主题的个人或是，如果是共同发明，全部个人”。在美国，要求保护发明的发明人或是共同发明人中的每个个人原则上必须对申请进行宣誓或作出声明。例如，在坦桑尼亚，“发明人”一词被认定为仅指人类，因此在现行法律框架下，仅由人工智能生成的发明不予承认——见坦桑尼亚答复 C.9304 的评论意见。

168 见菲律宾答复 C.9304 的评论意见，其中指出，根据菲律宾法律，人工智能系统不得被认定为自然人或法人，且缺乏持有、转让或行使知识产权所需的民事人格和法律行为能力；因此，在 IPOPHL 的申请中，人工智能不得被列为发明人、制造者、设计人或申请人，包括在共同发明或共同设计的情况下，相关人员同样必须是菲律宾法律承认的自然人；法国指出，迄今为止，法国上诉法院或最高法院尚无关于人工智能系统可能被指定为发明人的判例法；或者根据 INPI，发明人必须是自然人，且在申请提交时指定——见法国答复 C.9304 的评论意见。

169 专利、外观设计和商标总署，2025 年《计算机相关发明（CRI）审查指南》，第 5.2.1 节；印度专利局，申请号 202017019068，根据第 15 条做出的裁定（2026 年 4 月 15 日）。

170 见菲律宾答复 C.9304 的评论意见。

171 见美利坚合众国答复 C.9304 的评论意见。

213. 在当前做法下，尽管各司法管辖区的适用标准和门槛各不相同，但人工智能辅助的发明通常被视为在现有的发明人资格框架内。因此，发明人资格的确定需要依据相关法律评估一名或多名自然人是否对要求保护的发明做出了法律上充分的贡献。随着人工智能工具在发明过程中发挥越来越重要的作用，确定此类贡献的性质和程度，包括对人工智能系统开发、训练或运行的间接但针对具体发明的贡献，将变得日益复杂，可能需要进一步澄清。

214. 发明人指定不准确的法律后果因司法管辖区而异，可能取决于错误单纯涉及形式、是可更正的失误，还是权利归属的缺陷。如果申请人未指明发明人，专利局可能会请申请人在规定期限内更正该遗漏，否则申请可能被驳回。许多专利法还允许在审查过程中或授权后更正错误指定的发明人。在许多国家，发明人标注错误本身并不构成专利无效的理由。但是，如果不准确的指定表明申请人或专利权人并未从真正的发明人处取得权利，或者发明是通过侵占获得，则后果可能更为严重，包括专利无效或撤销、在涉及欺诈意图的情况下专利不可执行、将申请或专利转让给有资格的人、更正专利登记簿、损害赔偿或适用法律提供的其他救济。[172]

215. 一般来说，专利权在一开始属于发明人，而发明人可以把权利转让给其他自然人或法人。因此，发明人权利（即发明了发明的人）和所有权（即有权提交专利申请或获得专利的人）是两个相互独立但又相互联系的概念。在很多国家，如果发明属于职务发明，专利权原则上属于员工，对此通常附带某些条件。[173]因此，发明人权利/所有权问题可作为设计专利制度过程中关键政策问题的一部分。

216. 萨尔瓦多指出，其《人工智能与技术促进法》（2025 年 2 月 26 日第 234 号立法令）载有关于在萨尔瓦多境内开展的人工智能研究、开发或创新产生的知识产权的所有权和保护的具体规定。提交文件指出，所有此类知识产权，包括专利、版权、商标和商业秘密，均专属于开发实体或个人，而涉及多个实体的合作项目中的所有权则由各方事先达成的合同协议决定。该法还规定，在萨尔瓦多开发的专有算法、数据集和人工智能系统受法律保护，未经所有人明确同意不得公开或使用；国家应为获取适用于人工智能或类似技术的国际知识产权保护提供便利，包括协助开发人员根据 PCT 等全球条约注册创新成果。该提交文件未涉及人工智能系统是否可被列为发明人的独立问题。[174]

人工智能发明人权利和 DABUS 案

217. 随着人工智能系统在研发中的使用日益增加，专利制度不得不考虑当一项发明据称是由人工智能系统生成或在人工智能系统的辅助下生成时，关于发明人资格和权利归属的现有规定如何适用。这一问题有别于人工智能相关发明的可专利性。它涉及人工智能系统是否可被列为发明人，以及当所主张的发明贡献归属于人工智能时，申请或获得专利的权利应如何确立。

218. DABUS 申请已成为检验这些问题的主要测试案例。这些申请提交的依据是人工智能系统 DABUS 自主生成了相关发明，因此应被认定为发明人。随之而来的诉讼程序聚焦于门槛问题：根据现行专利法，“发明人”的概念是否能够延伸至非人类的机器。DABUS 案诉讼中出现的总体立场是目前尚无最终生效的实质性裁决接受人工智能系统根据现行专利法可被列为发明人。这一结论并不一定排除对人工智能辅助开发的发明给予专利保护，但强调了确定相关人类贡献并依据适用法律确立权利资格的持续重要性。

---

172 产权组织，《人工智能与发明人资格》（SCP/35/7 的更新），文件 SCP/37/5。

173 为了完整的目的，还应补充的一点是专利权也可以通过继承转让给他人。

174 见萨尔瓦多答复 C.9304 的评论意见。

219. 因此，为本研究之目的，DABUS 案最好理解为说明了在没有确定人类发明人的情况下，现行的发明人资格、发明人指定及权利资格要求如何运作。该案并未解决有关人工智能辅助发明的所有法律和政策问题，包括在将人工智能用作发明过程的一部分时，可能要求的人类贡献程度或性质。关于知识产权局和法院对 DABUS 案裁决的更详细说明在文件 SCP/38/5 中提供。

四、作为工具应用于专利制度的申请过程和管理工作的人工智能技术

220. 人工智能技术解决方案可被用于专利程序以及专利程序以外的情况中，即作为申请人提交专利申请的工具、专利局处理专利申请的工具、专利权人维权的工具、第三方提出宣告专利无效请求的工具或司法机关解决争议的工具等。

A. 知识产权主管部门的工具

221. 知识产权局已开始利用人工智能技术为知识产权管理和服务交付提供便利。产权组织知识产权局人工智能倡议索引[175]是一个在线门户，在该网站上按国家/领土和人工智能商业应用开列有人工智能技术在各知识产权局的使用情况。索引中的商业应用类别（是人工智能技术为知识产权局提供便利的主要业务领域）有：(i)数字化和流程自动化；(ii)专利审查管理；(iii)问询处服务；(iv)图像检索；(v)机器翻译；(vi)专利分类；(vii)专利现有技术检索；(viii)商标分类；(ix)版权登记；及(x)数据分析。

222. 在 2018 年 5 月 23 日至 25 日在日内瓦举行的产权组织 ICT 战略和人工智能促进知识产权管理知识产权局会议期间，会议主题之一是知识产权局如何已经利用以及可以利用人工智能和其他先进技术的应用。[176]会议讨论表明多个主管局在把人工智能的潜力用于知识产权管理系统方面取得了进展，并且显示各主管局希望持续交流有关人工智能的信息和经验，这么做还可避免尤其是进行重复性工作。[177]作为会议的一项跟进工作，产权组织为人工智能创建了一个专门网页[178]，并为讨论 ICT 战略和人工智能促进知识产权管理建立了电子论坛，仅对知识产权局所指定的专家开放。此外，产权组织标准委员会（标准委）成立了 ICT 战略和标准工作队，负责特别是对提交会议的建议进行审查。[179]

223. 最近，在 2026 年 4 月，产权组织推出了产权组织信通技术领导力对话（WILD），为知识产权局的信通技术负责人和高级管理官员提供了一个论坛，以交流实践经验，并探讨知识产权业务和服务交付的数字解决方案和新兴技术方面的合作。WILD 重点关注人工智能、网络安全和数据管理等领域，作为各知识产权局之间务实合作的自愿、非约束性平台运作。继首届会议后，除其他事项外，WILD 成立了人工智能工具协作小组，汇集来自各知识产权局的人工智能专家和产品所有人，协作开发、评价和实施用于知识产权管理的人工智能工具，协调试点项目并分享人工智能解决方案。

224. 对于专利审查，WILD 的公开材料显示，各知识产权局正在就人工智能辅助专利检索、专利分类、审查员及形式支持、现有技术识别以及相关工作流程工具交流经验。例如，WILD 第二届会议包括讨论利用人工智能支持审查员和形式审查人员、介绍基于人工智能的专利审查解决方案，以及就各知识产权局当前使用的人工智能工具和合作机会开展交流。因此，WILD 通过提供务实的

---

175 https://www.wipo.int/about-ip/en/artificial_intelligence/search.jsp。
176 会议的文件和演示文稿见：https://www.wipo.int/meetings/en/details.jsp?meeting_id=46586。
177 文件 WIPO/IP/ITAI/GE/18/5（协调员的总结）。
178 https://www.wipo.int/about-ip/en/artificial_intelligence/。
179 文件 CWS/6/3。

合作渠道与产权组织知识产权局人工智能倡议索引和 SCP 交流会相辅相成，主管局可通过该渠道比较审查相关的人工智能用例、经验教训及联合开发的可能性，同时将任何特定工具的采用留给每个主管局自行决定。[180]

225. 此外，自 2019 年 2 月以来，产权组织知识产权与人工智能对话（后更名为产权组织知识产权与前沿技术对话）提供了一个开放、包容的论坛，就人工智能等前沿技术对知识产权的影响，与尽可能广泛的利益攸关方进行接触，并促进他们之间的讨论和知识积累。[181]最近的会议继续探讨人工智能相关问题，包括人工智能发明、知识产权局对人工智能的使用以及生成式人工智能引发的问题。

226. 在专利管理领域，国家和地区专利局开发了（或正在开发）人工智能应用工具用于：专利申请分类；手续检查、现有技术检索；相关文件的机器翻译；协助实质审查（如自动标注专利文献以及自动检测不属于可授权主题的主题）；以及更为普遍的数据转换和文件管理。[182]

227. 产权组织国际局也在工作中使用人工智能，以强化本组织的职能和程序。国际局还开发并提供了一系列人工智能驱动的服务和工具，为用户和利益攸关方提供帮助。产权组织目前在多个领域使用并提供基于人工智能的工具和服务，包括语音转文字、全球品牌数据库内的图像相似性检索、通过 IPCCAT 的自动专利分类、通过全球商品和服务浏览器支持的尼斯分类、维也纳分类助手以及通过 WIPO Translate 的机器翻译。[183]

228. 在 SCP 第三十一届、第三十三届和第三十五届以及第三十六届会议的交流会上，各知识产权局和产权组织国际局介绍了它们在专利管理和专利审查中使用人工智能的情况。[184]SCP 第三十八届会议还包括一场分享会，涉及将包括人工智能在内的各种工具用于有效的专利审查程序，以及知识产权局面临的相关挑战。成员国还将分享他们在该领域新的合作途径方面的经验，秘书处也将介绍关于此类工具的举措。

B. 申请人、第三方和知识产权专业人员的工具

229. 申请人、专利权所有人、第三方和知识产权专业人员在与专利申请、审查、专利组合管理、执行和有效性评估相关的活动中，也越来越多地使用人工智能工具。这些使用不仅限于知识产权局开发的工具，还包括产权组织和各知识产权局公开提供的检索、分类和翻译工具，以及申请人、从业人员和专利体系其他用户使用的商业和通用人工智能系统。

230. 从功能角度看，此类工具可协助文档自动化、流程自动化以及人工智能实现的分析。举例而言，它们可支持现有技术检索、分类、翻译、校对、一致性检查、法律状态监控、专利态势、专利组合分析以及技术信息的整理或摘要。生成式人工智能和语义搜索工具的日益普及，可进一步协助用户制定检索策略、审阅大量专利和非专利文献、比较技术特征，并编拟供人工审查的初稿或工作文件。

---

180 产权组织，产权组织信通技术领导力对话（WILD）网页；产权组织信通技术领导力对话（WILD）第一届会议，2025 年 4 月 14 日至 16 日，日内瓦；产权组织信通技术领导力对话（WILD）第二届会议，2026 年 4 月 14 日至 16 日，日内瓦；产权组织，WILD/2/1，日程安排，2026 年 4 月 21 日。

181 https://www.wipo.int/about-ip/en/frontier_technologies/frontier_conversation.html。

182 产权组织知识产权局 AI 倡议索引。另见文件 SCP/32/4 和 4 Corr.以及 SCP/34/4；例如见葡萄牙答复 C.9304 的评论意见。

183 欲了解详细信息，请访问产权组织网站：https://www.wipo.int/en/web/ai-tools-services/。

184 演示文稿可在各 SCP 会议网页上查阅。文件 SCP/32/4 和 4 Corr.以及 SCP/34/4 分别载有 SCP/31 和 SCP/33 期间举行的交流会的报告。SCP 第三十六届会议期间也将举行类似的交流会。

231. 同时，应将人工智能工具视为对人类专业判断的辅助，而非替代。人工智能系统生成的输出可能不完整、不准确或缺乏依据，用户在专利相关事务中依赖此类工具时，可能需要考虑保密性、数据安全、职业责任以及任何适用的程序要求。

[后接附件]

产权组织和各知识产权局发布的有关人工智能的网页和出版物

人工智能与知识产权
https://www.wipo.int/zh/web/frontier-technologies/artificial-intelligence/index

知识产权和前沿技术
https://www.wipo.int/zh/web/frontier-technologies

人工智能工具和服务
https://www.wipo.int/zh/web/ai-tools-services

知识产权局人工智能举措索引
https://www.wipo.int/zh/web/ai-tools-services/ipos-initiatives

产权组织知识产权与人工智能对话会
https://www.wipo.int/zh/web/frontier-technologies/artificial-intelligence/conversation

人工智能基础设施交流（AIII）
https://www.wipo.int/en/web/frontier-technologies/artificial-intelligence/ai-infrastructure-interchange

帮助创新生态系统做好准备迎接人工智能-知识产权政策工具包（2024 年）
https://www.wipo.int/edocs/pubdocs/en/wipo-pub-2003-en-getting-the-innovation-ecosystem-ready-for-ai.pdf

2019 年产权组织技术趋势：人工智能
https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf

专利态势报告-生成式人工智能（2024 年）
https://www.wipo.int/web-publications/patent-landscape-report-generative-artificial-intelligence-genai/index.html

生成式人工智能：知识产权导航（2024 年）
https://www.wipo.int/edocs/pubdocs/en/wipo-pub-rn2024-8-en-generative-ai-navigating-intellectual-property.pdf

人工智能与发明人资格（SCP/35/7 的更新）
https://www.wipo.int/edocs/mdocs/scp/zh/scp_37/scp_37_5.pdf

学习机器——中小企业人工智能与知识产权入门
https://wipo-dpp.gateb.com/tempDownload/120DC6C1B13095019A0B0A57974086D223573983

澳大利亚

生成式人工智能与知识产权制度

https://www.ipaustralia.gov.au/temp/Generative-AI-and-the-IP-System.html

https://www.ipaustralia.gov.au/tools-and-research/professional-resources/data-research-and-reports/publications-and-reports/generative-ai-and-the-ip-system

专利手册，5.4.3 人工智能——发明人资格与权利归属
https://manuals.ipaustralia.gov.au/patent/5.4.3-inventions-produced-by-artificial-intelligence

巴西

金砖国家知识产权局最终报告，IP BRICS-INPI

http://www.ipbrics.net/secondpage/project/Patent%20Processes%20and%20Procedure%20-%20AI%20Study%20Report.pdf

加拿大

加拿大知识产权局，人工智能与知识产权
https://ised-isde.canada.ca/site/canadian-intellectual-property-office/en/artificial-intelligence-and-ip

中国

涉及计算机程序的发明专利申请审查
https://www.cnipa.gov.cn/art/2024/1/18/art_2199_189877.html

《关于修改〈专利审查指南〉的决定》（局令第 84 号）
https://www.cnipa.gov.cn/art/2025/11/13/art_74_202560.html?utm_source

法国

INPI, Le role du brevet dans les inventions utilisant l'intelligence artificielle
https://www.inpi.fr/actualites/dossiers-thematiques/role-du-brevet-dans-inventions-utilisant-lintelligence-artificielle

日本

“人工智能相关发明”网页

https://www.jpo.go.jp/e/system/patent/gaiyo/ai/index.html

人工智能相关发明的最新趋势（2026 年）
https://www.jpo.go.jp/e/system/patent/gaiyo/ai/ai_shutsugan_chosa.html

日本特许厅-中国国家知识产权局人工智能相关发明比较研究（2023 年）

https://www.jpo.go.jp/e/news/kokusai/cn/document/ai_report_2023_e/cn_ai_report_en.pdf

日本特许厅与欧专局计算机实施发明/软件相关发明比较研究（2021 年）

https://www.jpo.go.jp/e/system/laws/rule/guideline/patent/document/ai_jirei_e/01_en.pdf

与人工智能相关技术有关的专利审查案例示例
https://www.jpo.go.jp/e/system/laws/rule/guideline/patent/ai_jirei_e.html'

菲律宾

IPOPHL，人工智能相关发明、实用新型和外观设计申请的审查指南
https://www.ipophil.gov.ph/services/patent-examination-guidelines/

大韩民国

https://www.kipo.go.kr/upload/en/download/Examination_Guide.pdfKIPO，人工智能和可专利性 https://www.kipo.go.kr/en/HtmlApp?c=92012&catmenu=ek03_08_05

KIPO，按技术领域分列的审查做法指南——第一章，人工智能领域（2026 年）
https://www.kipo.go.kr/upload/en/download/EXAMINATION_PRACTICE_GUIDELINES_BY_TECHNOLOGY_FIELD.pdf

美利坚合众国

“人工智能”网页
https://www.uspto.gov/initiatives/artificial-intelligence

人工智能相关资源
https://www.uspto.gov/initiatives/artificial-intelligence/artificial-intelligence-resources

欧洲专利局

“人工智能”网页
https://www.epo.org/en/news-events/in-focus/ict/artificial-intelligence

欧专局审查指南，G-II，3.3.1 人工智能与机器学习（2026 年）
https://www.epo.org/en/legal/guidelines-epc/2026/g_ii_3_3_1.html

[附件和文件完]