

当面交付
香港添马
添美道 2 号
政府总部西翼 23 楼
商务及经济发展局
第 3 部

2024 年 9 月 20 日

敬启者：

版权与人工智能公众咨询意见书

我们就商务及经济发展局及知识产权署的版权与人工智能公众咨询文件（下称「咨询文件」）提交意见。

请注意，我们不愿意公开我方的姓名及背景资料，或是直接在其他文件中引述有关我方的任何背景资料。

1. 人工智能生成作品的版权保护

我们同意政府的看法，即香港法例第 528 章《版权条例》中有关电脑产生作品的现有条文，其涵盖范围已有足够弹性，足以容纳不断演进的科技，并将适用于人工智能生成作品。有关人工智能生成作品的版权存在问题，以及该等作品的作者身分问题，应根据创作该等作品生成过程的实际情况而定。特别是由于人工智能技术仍在快速发展之中，在现阶段对法例进行任何修改都为时过早。

2. 人工智能生成作品侵权的法律责任

关于人工智能生成作品可能侵权的法律责任，我们同意现时没有必要修改现行法例。AI 技术现时在某种程度上仍是个「黑盒子」，因此 AI 系统最终能够生成什么样的作品并不具明确的可预测性，加上每个 AI 模型的运作亦可能有差别，且 AI 开发方与最终使用者在具体作品的生成过程中所发挥的具体作用，亦因场景的不同而存在差异，所以我们认为在现阶段以修例方式一刀切硬性规定侵权法律责任特定地由某人（例如人工智能系统的拥有人及 / 或最终使用者）承担的立法方法并不适合。因此，我方认同侵权的法律责任需要按个别情况来决定，并须顾及每名有关人士在侵权行为中的个别角色和参与程度。

3. 拟议文本及数据开采豁免

我们注意到香港政府有意将人工智能融入各行各业及推动人工智能的广泛应用。¹ 尽管我们相信这将增加 AI 模型提供者未来在香港的商机，但目前 AI 产业尚处发展初期，需要大量且持续

¹ 如：创新科技及工业局局长孙东教授在 2024 年 1 月 31 日的立法会会议上表示，政府透过多项措施及资助计划，协助不同产业透过科技升级转型，例如推出「科技券」计划支援本地企业 / 机构使用科技服务和方案（包括人工智能技术）。（见 <https://www.info.gov.hk/gia/general/202401/31/P2024013100367.htm?fontSize=1>）

的经济及人力投入。我们认为全球绝大多数 AI 开发方还远不具备盈利能力，甚至于 AI 产业可能的盈利方向、盈利模式在现阶段均尚不明晰。同时，AI 产业还面临着算法研发、算力芯片开发等硬性成本。目前讨论 AI 产业相关法律、政策的前提，我们认为并敦促政府考虑尽可能为 AI 产业的发展降低成本负担、提供充足的发展空间，以及提供对于研发所必备的版权数据的公平合理的交易机会。这亦符合国家以及香港政府有意推进香港科研发展及人工智能开发的整体规划格局及方针。

我们支持引入拟议文本及数据开采豁免，认为有助 AI 产业发展，详细理由如下：

3.1 豁免范围

首先，我们认为豁免应适用于：

- (i) 传统的文本及数据开采，即采用自动化技术提取大量文本、图像、数据及／或其他类型的资料，并进行运算分析，以得出未必可单从人手分析取得而有价值的见解、模式、趋势和关联；及
- (ii) 对大量文本、图像、数据及／或其他类型的资料进行运算分析和处理，以提升电脑程式的效能，当中包括开发、训练和提升人工智能模型，尤其是生成式人工智能模型。

我方亦认同及支持拟议文本及数据开采豁免有必要同时适用于商业目的和非商业目的的数据挖掘和 AI 训练。从 AI 产业的经济成本角度看，如若将豁免限制于非商业使用场景，将导致商用 AI 模型在训练阶段面临高额的许可费成本和许可被随时中止的风险，这对于吸引市场主体或国际投资者在 AI 产业进行商业投入毫无助益，也难以发挥推动 AI 产业发展的作用。如将豁免限制于非商业用途，将有违香港政府引入豁免的原本目的。

此外，无论政府是否会授予版权权利人「选择退出」豁免的权利（见下文第 3.3 节），我们认为拟议文本及数据开采豁免应具有追溯力。如下文第 3.3 节所述，由于有关的版权内容、数据及相关训练内容可能在拟议文本及数据开采豁免生效前已融入于 AI 模型中，从 AI 模型排除版权内容在实践中难以实现。因此，豁免应适用于将来及豁免生效前已用于训练 AI 模型的版权作品及数据。

3.2 引入拟议文本及数据开采豁免的必要性

简单而言，在训练大部分的模型过程中，一般 AI 开发方通常都有需要在某一阶段于 AI 系统的神经网络架构上复制并储存大量的数据（其中不可避免地包含有版权作品）然后进行分析。由于现行技术一般未能容许神经网络在只抓取而不储存作品的情况下进行分析，因此，一般而言，复制及储存作品在某些模型（如大型语言模型）的训练过程通常是无法避免的。

在上述前提下，我们支持引入同时适用于商业目的和非商业目的的拟议文本及数据开采豁免，这将为 AI 开发方提供法律上的明确性及减低开发模型经济成本，从而促进人工智能及香港科研的发展。

虽然 AI 开发方理论上可以向版权拥有人索取特许以使用版权作品进行训练，但现实上，版权许可交易的情况现时不普遍，在现实情况下也难以操作。我方认为如要求 AI 开发方取得版权拥有人就使用其版权作品训练人工智能模型，将使 AI 开发方陷入重重困难，严重窒碍人工智能发展：

- (a) 从模型研发（训练数据的充足性）角度，取得全球范围内绝大多数版权权利人的书面授权在实践层面不具备可操作性，其原因包括：
- 版权作品和作者数量巨大：版权作品种类繁多，权利人过于分散，而 AI 模型训练需要大量高质量数据，实践中难以通过逐一取得授权的方式收集到足够的数据以供训练，或者说，上述方式会极大、极不合理地加剧 AI 模型方面临的成本；
 - 版权作品权属/授权流转情况庞杂：版权作品存在授权链条和分散授权情形，例如说：版权拥有人在进行版权许可时，可能限制了授权的使用时间、使用地区以及 / 或版权作品的用途限制，版权拥有人的被许可人不一定有充分的许可再授权给 AI 开发者；而同一版权作品在同一时间 及/或同一地域亦可能有不同的被许可人，逐一核实大量类型各异的版权作品的权利人及/或其权利范围极其困难，这将使 AI 开发方负担过重。举例而言，版权在香港是不可注册的，也没有中央及/或官方的版权资料库。因此，在进行文本及数据开采活动之前，对版权拥有权逐一进行尽职调查并向相关版权权利人寻求授权在实践中难以操作；及
 - 现有集体管理组织不足以涵盖所有版权作品：现时香港的注册特许机构为：(1) 香港作曲家及作词家协会有限公司 (CASH)、(2) 香港音像联盟有限公司、(3) Motion Picture Licensing Company (Hong Kong) Limited、(4) 香港音像版权有限公司、(5) 香港复印授权协会有限公司、及 (6) 香港书刊版权授权协会有限公司。许多用于训练模型的作品(如：网页、网志等)不被涵盖于上述的注册特许机构，因此，仍会有大量版权作品无法找到正确的权利人。
- (b) 从商业成本角度，AI 开发方与版权方关于授权的谈判时间成本极高，在实践中授权费用往往难以达成一致，对于 AI 开发方来说经济成本极高且存在不确定性。现阶段，用于大模型开发的数据（包括版权内容）本身的价值无法通过市场得以展示或定价、从购买数据到训练后形成稳定的商业模式需要比较长的时间，因此我方认为版权权利人对授权价格的预期不应过高，但同时相当多版权权利人或语料库企业对其持有的结构化数据都抱有乐观态度，他们宁愿待价而沽或者在谈判时“敲竹杠”。我方也留意到理性的版权权利人已经认识到即使授权给 AI 开发方，也未必能得到即时回报，考虑到市场的不确定性，更有可能抬高价格，赚取短期收益。特别是那些已经拥有广泛内容生态的平台企业、语料库企业会更加谨慎，因为其无法测度内容授权共享给 AI 开发方带来的间接流量收益，更担心 AI 开发方将通过不断学习优质数据而成为内容生产者或生产链条上游的工具垄断者，降低整个内容服务市场的成本并占据竞争优势。在双方不同预期下，谈判很容易破裂。即使 AI 开发方有意通过支付合理许可费用的方式取得授权，由于缺少相应的制度规制，难以找到所有版权方，亦难以促成相关许可交易。此外，与每个版权权利人分别协商将非常耗时，因此严重妨碍模型研发的进展。
- (c) 从 AI 产业与版权行业竞争业态的角度，我方认为版权方一般对人工智能产业天然存在警惕，在实践中谈判难度高，交易难度亦大，如要求 AI 开发方取得版权拥有人就使用其版权作品训练人工智能模型的授权，这将使 AI 产业在发展初期便立即面临的经济成本加剧，这亦不利于 AI 产业发展和国际竞争力的增强。由于高质量版权内容的不可替

代性，如版权方在谈判时坚持极高且不合理的授权费用，AI 开发方在授权谈判中势必处于被动地位，导致许可费用高昂、许可交易难以实现，导致在实践中，就文本及数据开采活动进行版权许可交易的情况少有实际存在。

- (d) 从 AI 产业发展的稳定性角度，版权特许交易的商业谈判不仅存在经济成本、时间成本，而且具有不确定性：即使 AI 开发方获得许可使用版权内容作训练模型用途，仍然可能导致 AI 开发方在许可期限届满后，面临前期训练成果无法投入实际使用的风险。对于已经投入使用的 AI 模型产品需要排除许可到期或届满的版权内容，重新训练和数据遗忘的成本极高，且由于有关的版权内容、数据及相关训练内容已经融入于 AI 模型中，再加上不同的版权许可定必有不同的许可期限及到期日，从 AI 模型排除版权内容在实践中难以实现。此外，即使 AI 开发方重新训练 AI 模型，在实践中排除到期版权内容或资料也可能会影响模型的准确性，且移除版权内容或资料也可能会导致模型效能降低，尤其是当被删除的资料具有重要的预测能力时。

因此，即使 AI 开发方与版权权利人续展版权许可，鉴于从 AI 模型中排除版权内容的困难和成本，在续约过程中，版权权利人仍将处于绝对的谈判优势地位，这很可能导致授权费用不合理持续增高，令 AI 开发方持续处于不合理的极其被动的谈判地位。

- (e) 从增加区位优势的角度，就 AI 发展及应用，全球立法存在差异，目前已有部分国家和地区引入文本及数据开采豁免制度，产业资源和上下游企业很可能逐渐向上述地区倾斜，如仍然要求对版权内容的训练行为需要取得权利人相关授权，明显不利于 AI 行业在香港发展。这窒碍了人工智能发展以及大幅削弱香港在科研发展中的竞争力。

总括而言，我们认为单靠 AI 开发方取得版权拥有人的特许对训练人工智能模型、文本及数据开采活动并不可行。我们相信，引入拟议文本及数据开采豁免将可减少与版权权利人协商授权安排的成本和时间，从而更有效地促进人工智能的发展。

3.3 拟议文本及数据开采豁免的条件

咨询文件中建议对拟议文本及数据开采豁免施加一定条件，包括：

- (a) 使用者（如 AI 开发方）必须合法取用版权作品（下称「**拟议条件(a)**」）；
- (b) 如版权作品有特许计划可供应用，或版权拥有人已表明保留其权利（即「选择退出」选项），则相关文本及数据开采活动即属未经授权（下称「**拟议条件(b)**」）；及／或
- (c) 就进一步传播／分发／经销在拟议文本及数据开采豁免下制作的复制品施加限制（下称「**拟议条件(c)**」）。

从区域竞争和产业发展角度看，我们认为拟议文本及数据开采豁免不宜附带过多限制。我们将先对拟议条件(b)作回应，然后再阐述我们对拟议条件 (a) 和 (c)的看法：

拟议条件 (b)：如版权作品有特许计划可供应用，或版权拥有人已表明保留其权利(即「选择退出」选项)，则相关文本及数据开采活动即属未经授权 — 我们认为不应赋予版权权利人以单方通知形式「选择退出」的权利

以欧盟的做法为例，我们认为从欧盟的实践中可见欧盟的「选择退出」机制不利于人工智能产业的发展。另一方面，新加坡目前的法例是不允许版权权利人以合约方式凌驾法定的文本及数据开采豁免，版权权利人也没有「选择退出」选项。由于新加坡亦正考虑进一步放宽文本及数据开采活动可获豁免的条件（见下文），我们认为香港应考虑邻近司法管辖区的做法，以保持其作为人工智能发展枢纽的竞争优势。但假如香港最终决定引入「选择退出」选项（我们并不支持），我们认为版权权利人在单方面行使「选择退出」时必需承担一定的经济成本（详见下文），否则，拟议文本及数据开采豁免将不能达到香港政府原本引入豁免的目的。

欧盟的做法

我们理解欧盟第 2019/790 号《数码单一市场版权及相关权利指令》（下称「**欧盟指令**」）的豁免仅适用于版权拥有人未有以适当方式（例如就网上公开的内容采取机器可读方式）表明保留其版权作品及其他标的物的权利的情况。结合欧盟等地的实践，我们认为，赋予版权权利人以单方通知形式「选择退出」的权利，将导致版权权利人可在没有任何成本的情况下，从上游环节遏制 AI 产业的发展。在此情况下，「选择退出」一定程度上将退化成版权权利人谋求超额垄断利益的工具。此外，AI 开发方在实践中不可能，或需花大量时间及资源检查每个网站的条款和细则，查阅权利人是否允许抓取内容作模型训练用途。此外，由于有关的版权内容、数据及相关训练内容可能在版权方「选择退出」之前，已在训练时融入于 AI 模型中，再加上不同版权方「选择退出」的时间点有异，从 AI 模型排除版权内容在实践中难以实现，因此「选择退出」机制无法解决以上第 3.2 节所述的 AI 开发方的重重困难。正是基于上述原因，欧盟的「选择退出」机制才被普遍认为形同虚设，广受诟病。

AI 产业的发展需要高质量数据的收集与汇聚，而不是分散的垄断性使用。如大量知名版权方均有机会采取随时退出的策略，由于版权内容本身存在分散性，即使大版权方出于利益诉求希望使用其版权内容自行开发 AI 产品，亦会出现任何版权方/模型方都无法获取/汇集足够数量的高质量版权数据、而仅能基于自身持有的版权内容分别进行训练的情形，高质量数据的分散与缺乏也将直接影响模型质量与技术进步，进而影响 AI 产业发展。

新加坡的做法

反之，新加坡《2021 年版权法令》禁止版权权利人以合约凌驾文本及数据开采豁免。我们也了解新加坡政府于 2024 年 4 月展开公众咨询，就版权作品使用者（如 AI 开发方）是否可以应用规避控制存取版权作品的科技措施的情况，以进行文本及数据开采征询意见。²在公众咨询中，新加坡政府就禁止规避存取控制措施是否可能损害或不利地影响新加坡《2021 年版权法令》的文本及数据开采豁免范围内允许使用作品进行计算数据分析征求意见。因此，新加坡的文本及数据开采豁免可能会进一步放宽。如若香港法律规定的拟议文本及数据开采豁免较之邻近区位的竞争对手更加严格，这将不利于 AI 产业向香港靠拢、聚集。

因此，我们认为不应赋予版权权利人以单方通知形式「选择退出」的权利。

² 見 <https://www.mlaw.gov.sg/public-consultation-on-ca-exceptions2024/>

但假如香港最终决定引入「选择退出」选项（我们并不支持），版权权利人与 AI 开发者可通过有效协议约定的方式排除数据挖掘，但版权权利人单方面行使「选择退出」权利时应设置经济成本。

我们注意到，香港立法机关对于以市场方式解决 AI 开发中的数据挖掘问题，始终保持关注。但我们也认为，在立法层面有必要对于缔约双方的市场地位予以充分、必要的关注。从目前的市场格局出发，版权方作为掌握训练数据并且基于法定权利垄断该部分训练数据的主体，在商业谈判中较之 AI 模型方具有天然的谈判优势。事实上，从目前的情况来看，版权方基于这种谈判优势直接拒绝谈判的情况也并不鲜见。在此情形下，引入「协议约定方式排除数据挖掘的前提，是我们希望政府坚持为版权权利人单方面「选择退出」设置必要的经济成本。我方相信只有在权利方单方面选择退出而需要付出一定的经济成本时，才有可能就版权内容许可事宜产生动力与 AI 开发方进行接洽并启动谈判，并出于规避相应经济成本的目的，尝试与 AI 开发方通过商业安排，避免其作品成为数据挖掘的对象。因此，为版权方「选择退出」设置必要成本的目的，一方面在于敦促版权方将其版权内容投入市场中，与 AI 开发方进行公平合理的谈判，促进作品的使用与流转；另一方面也为 AI 开发方创设基本的谈判机会，使其有可能通过支付相对合理的对价，获取模型开发所必需的版权数据。

同时，对于传统的版权许可交易（传统许可范围，一般不包括以使用作品进行数据训练为目的的授权交易），版权权利人亦可通过谈判，就在许可范围内排除数据训练事宜，与权利人达成合意及协议条款，或对数据训练的授权许可提出单独的合理对价。

从促进 AI 产业长期稳定发展的角度看，假设香港法律在规规定拟议文本及数据开采豁免的同时赋予版权权利人随时「选择退出」机制，由于权利人行使「选择退出」权利的时间节点具有不确定性，如在香港 AI 产业兴盛之际，出现权利人集中「选择退出」的情形，也会对 AI 产业发展造成致命打击；前述风险的存在，也必然会反过来降低市场主体及投资者在香港投资 AI 产业的信心。

我们明白需要衡量版权权利人，但认为必须对版权权利人单方面「选择退出」设置必要的经济成本，以在保护权利人权益的同时，为 AI 产业发展保留必要的空间。举例而言，必要的经济成本可以是：

(i) 版权权利人证明其已经采取了合理且有效的技术措施，防止其作品被抓取；及

我们也考虑过让版权权利人拥有单方「选择退出」的权利，但必要要求版权权利人证明版权权利人已经采取了合理且有效的技术措施，防止其作品被抓取。

我们认为，当版权权利人不是以单方通知的形式「选择退出」，而是投入一定程度及规模的人力、物力、财力开发技术措施，防止其作品被模型方自行抓取时，相对有理由相信版权权利人认为其作品本身较之一般作品更具有价值、属于版权权利人较为核心的资产。在此情况下，允许版权权利人通过承担合理程度的成本进而「选择退出」，具有一定合理性。

(ii) 版权权利人已明确存在在相关市场将自身作品许可用于 AI 训练的商业行为 / 实际意图，且版权权利人需要向 AI 开发方明确表示「选择退出」的意图，而权利人亦需得到 AI 开发方的确认

若香港要引入「选择退出」的权利，我们认为版权权利人必须已明确存在在相关市场将自身作品许可用于 AI 训练的商业行为 / 实际意图，且版权权利人需要向 AI 开发方明确表示「选择退出」的意图，而权利人需得到 AI 开发方的确认，才能有效地「选择退出」。

这是因为，如果版权拥有人本身也不打算就版权作品进行任何有关 AI 训练的商业许可行为，禁止 AI 开发商使用该等版权作品训练 AI 模型会使该等作品被「闲置」，窒碍市场和经济的发展。我们认为，当版权权利人已明确存在在相关市场将自身作品许可用于 AI 训练的商业行为 / 实际意图时，即使其作品仅被有限的 AI 开发者使用，至少其作品未被「闲置」，仍可能为香港 AI 产业的发展发挥促进作用。在此情形下，允许版权权利人利用其对作品享有的专有权获得合理收益，为其 AI 产品的开发创造一定竞争优势，促进数据的自由流动和交易，亦具合理性。

考虑到版权人容易通过签订阴阳合同等形式伪造实际使用意图，我们还认为有必要对“商业行为”和“实际意图”赋予必要要件 (例如：有关的“商业行为”和“实际意图”必须是有偿的公平交易，而且并不能以一些安排和计划将该等有偿代价以任何形式归还或转移至版权权利人)，以确保版权人已经或者即将投入实际成本，开始进行相关运营。

同时，我们认为政府亦需注意通过立法排除版权权利人虽在形式上存在许可作品的意图，但实质上未在相关市场进行任何商业行动或坚持远高于合理定价的高额授权费用 (或是立法设立独立的、程序简便的裁判机关决定合理授权费用)，从而规避上述第 3.2 节所述的情形。

「选择退出」机制的追溯力

在上述前提下，我们认为「选择退出」机制仅可约束其生效后的作品使用行为，不应溯及既往。我们认为，关于「选择退出」机制的适用范围，除需包含上述前提，还应关注版权权利人「选择退出」的时点：对于在「选择退出」机制生效之前已经开始进行或完成的数据挖掘和 AI 训练，仍使用拟议文本及数据开采豁免，不应再赋予版权权利人向 AI 开发者主张版权的权利。由于版权权利人数量巨大，且存在实体作品和数据作品等各种类型，如果「选择退出」意味着该权利人的作品自始不能被用于训练，AI 模型方将面临不定期重置数据集、重新训练及数据遗忘的高额成本，给 AI 产业发展造成难以避免的冲击，并将使拟议文本及数据开采豁免的机制因缺乏稳定性而无法在实践中适用，名存实亡。我们建议，对于版权权利人「选择退出」之前已经进行的数据挖掘和 AI 训练，应当适用一般性的、没有任何规限的文本及数据开采豁免完全禁止权利人向 AI 开发者主张权利，从而保 AI 证模型方可以在相对稳定的外部环境下发展产业。

拟议条件 (a)：使用者 (如 AI 开发方) 必须合法取用版权作品

我们原则上同意，使用者 (如 AI 开发方) 应合法取用版权作品才能享用豁免。但我们认为法例应该明确规定何谓「合法取用」。我们认为，考虑到数据挖掘过程中涉及巨量数据且相应挖掘系通过技术方式实现，「合法取用」的标准不宜设定过高。如若相应数据挖掘以训练 AI 模型为唯一目的而未用于其他受到《版权条例》规制的用途，且 AI 模型方是基于善意获取并使用数据，就应被认为是「合法取用」。例如，在 AI 模型方以合理对价从第三方获得数据包的情况下，鉴于数据包涉及巨量数据，而由于 AI 模型方在实际操作中

难以甚至无法逐一核实该等数据包的所有数据是否完全没有包含侵权内容，因此，即使第三方提供的数据包中包含少量侵权内容，也不宜直接将 AI 模型方的行为认定为非法取用。

拟议条件 (c)：就进一步传播/分发/经销在拟议文本及数据开采豁免下制作的复制品施加限制

由于识别人工智能生成作品的技术措施仍未成熟，因此可能难以识别此类作品，令限制传播/分发/经销人工智能生成作品的条件难以执行。此外，人工智能工具在生成内容时，必然会依赖根据拟议文本及数据开采豁免使用版权作品训练的模型，因此必然会涉及作品复制品的「进一步」交易。因此，我们认为引入此类条件将违背引入拟议文本及数据开采豁免的目的。

4. 对于版权权利人自身利益的维护与平衡，应当在生成阶段予以解决，不宜在训练阶段体现。

关于版权权利人对 AI 产业可能严重损害其自身利益的普遍担忧，我们认为，应当将模型开发的训练阶段，与内容生成的应用阶段分别单独进行考量。如果对版权权利人合法权益可能构成较大影响甚至损害的是 AI 产品最终生成的内容，而非技术或产业本身，则相关的利益平衡应当在内容生成阶段而非训练阶段予以体现，即：通过新技术以更高的效率创造出更多具有独创性的优质作品，同时保证原权利人的合法权益，而非在训练阶段通过数据垄断中止开发，导致技术发展过程中不可避免地在数据获取、使用及训练等层面出现混乱无序的情形，如此方可实质性地平衡各方利益，促进公平交易，兼顾产业发展。

具体而言，在生成阶段，AI 模型生成内容如与在先作品构成实质性相似，版权权利人有权提起侵权之诉，通过《版权条例》现有规定维护自身权益，向侵权方主张停止侵权及经济赔偿；而在训练阶段，由于版权权利人具有不可替代的数据垄断优势，在考虑到 AI 开发方寻求绝大多数权利人逐一授权的不可操作性，从而设置拟议文本及数据开采豁免的同时，亦为版权权利人提供了在相关市场自由交易版权数据、实施有偿许可的可能性，从而基于其持有作品获取合理利益的可能性已经对版权权利人基于其专有权享有的利益进行了充分有效的保障和平衡，同时确保高质量的版权数据能够在市场流通，保障 AI 技术进步和产业发展。