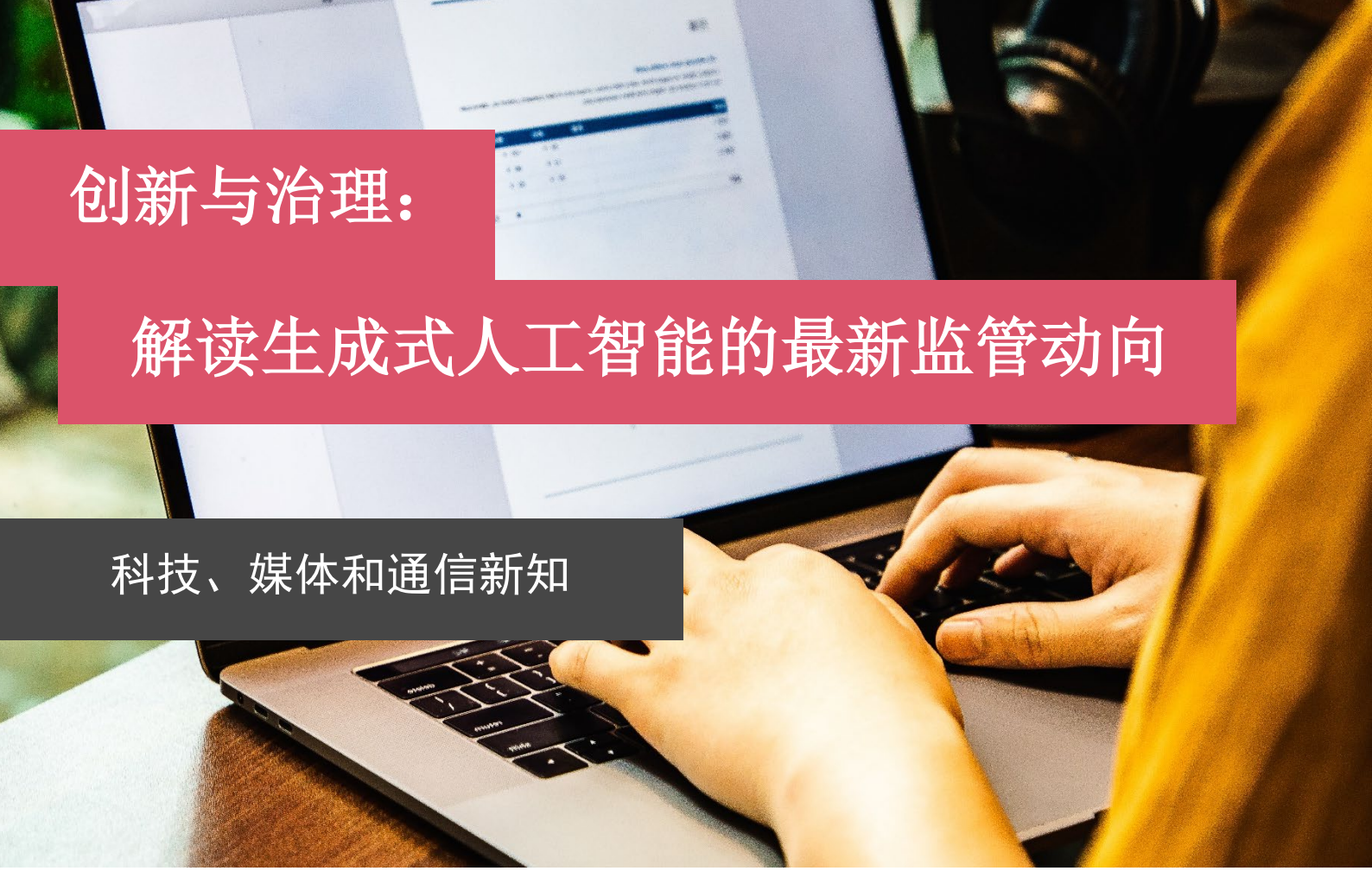


创新与治理：解读生成式人工智能的最新监管动向

科技、媒体和通信新知

2023年4月

简介

随着大数据时代的到来和计算机能力的迭代，人工智能（AI）技术在近些年进入了黄金发展期，各国对AI技术也开始施行不同程度上的监管。欧盟于2021年12月发布了《人工智能法案草案》，而该草案规定了对高风险AI系统的监管措施。根据该法规，高风险AI系统需要经过严格的评估和认证，以确保其符合欧洲联盟的法律和道德标准。美国虽然尚未颁布专门针对AI监管的全国性法律，然而，一些州开始在AI领域推出相关立法。在2022年间，包括加州在内的17个州对AI领域启动了立法进程。

人工智能研发公司OpenAI在2022年11月推出大语言模型（Large Language Model, LLM）应用ChatGPT，AI技术进一步在普罗大众前褪去神秘的面纱。生成式人工智能（AI Generated Content, 'AIGC'）的概念应运而生，也标志AI技术正式迈入通用人工智能时代。虽然人工智能产品极大地提高了工作和生活效率，但随着近年来一系列AI应用相关负面案例的发生，加剧了立法者对AIGC应用的担忧。

在尚未确定ChatGPT是否将成为打开“潘多拉魔盒”的那把钥匙前，对AIGC的监管成了众矢之的。意大利数据保护局于今年3月31日宣布暂时停止使用ChatGPT，并已就ChatGPT涉嫌违反数据收集规则展开调查。除此之外，法国国家信息自由委员会（CNIL）以及西班牙数据保护机构也决定对ChatGPT立案并展开调查。在欧盟层面，欧洲数据保护委员会（EDPB）于4月13日表示，其正在成立一个特别工作组，帮助欧盟各国应对ChatGPT，促进欧盟各国之间的合作，并就数据保护机构可能采取的执法行动交换信息。而美国商务部于4月11日表示，其正在就AI系统的潜在问责措施征求公众意见。

2023年4月11日，中华人民共和国国家互联网信息办公室发布了《生成式人工智能服务管理办法（征求意见稿）》（“**征求意见稿**”），拟对生成式人工智能在我国的开发及应用进行规范。该稿征求意见将于2023年5月10日结束。本文提取征求意见稿中的重点内容进行总结与分析。

1. “生成式人工智能”的定义

欧盟的《人工智能法案草案》（“**草案**”）将“人工智能”定义为“*使用一种或多种技术和方法开发的软件，基于给定的一组人类定义的目标，生成诸如内容、预测，影响与其互动的环境的建议或决策之类的输出*”。从字面上看，该草案规定的AIGC生成基础并不包括模型、规则等，因此引发了一些基础模型（包括LLM）是否应纳入监管的广泛讨论。

而根据《征求意见稿》第二条规定，生成式人工智能是指“*基于算法、模型、规则生成文本、图片、声音、视频、代码等内容的技术*”。这一定义与《互联网信息服务深度合成管理规定》关于“深度合成技术”的解释一脉相承，但征求意见稿进一步扩大并明确了生成式人工智能产品的定义范围，其不仅包括基于算法生成的技术，还应包括基于模型、规则生成的相关内容。因此，ChatGPT、Bard、Midjourney、剪映等软件，均被纳入《征求意见稿》所规定的生成式人工智能产品的范畴。

2. 适用范围

《征求意见稿》的适用范围以服务对象作为依据，只要是“面向中华人民共和国境内公众提供服务”，无论该提供者是境内还是境外主体，也无论该产品是否在境内研发或者使用，都需要符合征求意见稿的相关要求。因此，如同《个人信息保护法》，其对AIGC的监管具有域外效力。

从字面上看，《个人信息保护法》对“域外效力”的规定似乎比《征求意见稿》具有更明确的定义；即若信息处理活动“*以向境内自然人提供产品或服务为目的*”或“*分析、评估境内自然人行为的*”，均受《个人信息保护法》管辖。然而，《征求意见稿》所指的“*面向中华人民共和国境内公众提供服务*”的边界存在一定的模糊，或将会产生如下问题：

i. 一些生成式人工智能产品（如ChatGPT）在客观上虽在全世界范围内可用，但在注册时，其需要用户提供相关的本地信息（比如当地手机号码或需要当地的信用卡）进行验证。若该等产品并不接受中国手机用户的注册，则该产品是否仍符合“面向境内公众”的要求？

ii. 另一些全球化的AIGC产品，其未对用户的国籍或所在地进行主观瞄准或筛选。即其从主观上或许并未设想“中国”为其目标市场之一。在网络无国界的假设下，此类产品是否属于《征求意见稿》所适用的“面向中国”的产品？

iii. 《征求意见稿》似表明其所调整的AIGC产品的服务对象须为“公众”。那么，如果某AIGC的服务并不针对公众，而仅针对某类特定客户，《征求意见稿》是否仍得适用？

鉴于人工智能技术正在以井喷式的速度发展，且考虑到人工智能技术将对社会产生重大的影响，在法律监管上，我们认为，就人工智能领域，立法者应在立法技术上做一定的弹性安排。综合考虑《网络安全法》、《数据安全法》、《个人信息保护法》以及国家对各类数据处理行为的约束态度，我们倾向于认为《征求意见稿》的适用范围应当作相应的宽泛解释，即无论生成式人工智能产品是否直接（主动式）或间接面向（被动式）中华人民共和国境内公众，都应当符合征求意见稿的规定，且“公众”亦应做一定的拓展性解释。

3. AIGC产品主要合规要求

《征求意见稿》对AIGC产品提出了若干合规要求，主要体现在数据安全、内容合规和知识产权保护三方面。

i. **数据安全：**AIGC系统通常需要大量的数据进行训练，包括图像、声音、文本等。这些数据可能包含用户的个人信息，如姓名、照片、声音录音等，如果未经充分保护，可能会导致用户的数据隐私泄露。同时，AIGC系统生成的内容可能包含用户的个人信息，如合成的人声、仿声、人脸生成、人脸替换等。这些生成的内容可能被用于不当用途，例如恶意识别、冒充他人身份、造假等，从而侵犯用户的隐私权。因此，《征求意见稿》规定了“服务提供者”需要履行个人信息保护义务，禁止非法获取、披露、利用隐私和商业秘密，不得随意进行用户画像等。这意味着AIGC产品在处理数据时需要确保数据的安全性，合法合规地使用用户的个人信息，并不滥用用户数据进行画像等行为。

ii. **内容合规：**《征求意见稿》规定了AIGC生成的内容应当真实、准确、客观、多样，并体现社会主义核心价值观，不得带有歧视性，并要求提供者采取措施防止生成虚假信息，因此该要求引起了不小的争议。首先，从内容来源来看，提供者数据库中的信息获取来源复杂且难以追溯，在模型训练时或可能进一步采用衍生信息，因此难以对数据库所有信息的真实准确性进行事前掌握。另一方面，由于人工智能生成的内容本身就具有一定的“创造性”，很难有单一的标准来判断生成内容是否“真实准确”。对此，还需要进一步的技术细则予以明确，或者可以通过其他方式，例如强化对来源难以追溯、创新内容等生成的信息进行特殊标识，进行事前警示。

iii. **知识产权保护**：《征求意见稿》规定了在数据预训练、优化训练和产品服务提供过程中，AIGC产品不得侵犯知识产权。AIGC在开发及应用过程中的知识产权问题也引发了热议。这其中涉及以下两个个方面：

- 训练数据的知识产权侵权：在AIGC系统的训练过程中，可能使用了大量的数据和模型，涉及多方之间的知识产权的安排，否则将极易构成侵权。例如，近期某图库供应商起诉某AIGC产品，指控后者滥用超过1200万张其照片来训练其Stable Diffusion AI图像生成系统。
- AIGC产品的版权归属：目前各国法律一般对纯人工智能生成的内容不给与知识产权保护。美国版权局在其取消漫画书《黎明的扎利亚》著作权登记时表示“非人类作者创作的作品无法进行著作权登记”。中国的《著作权法》亦有类似的规定。但AIGC是否存在著作权问题较为复杂，需要具体分析。如就Dreamwriter案例[1]，法院虽然不给与人工智能生成的内容保护，但给与包含编辑团队、产品团队、技术开发团队在内的主创团队运用Dreamwriter软件完成的涉案文章的著作权保护。因此，企业需要仔细周密地谋划如何妥善保护与人工智能作品相关的知识产权。

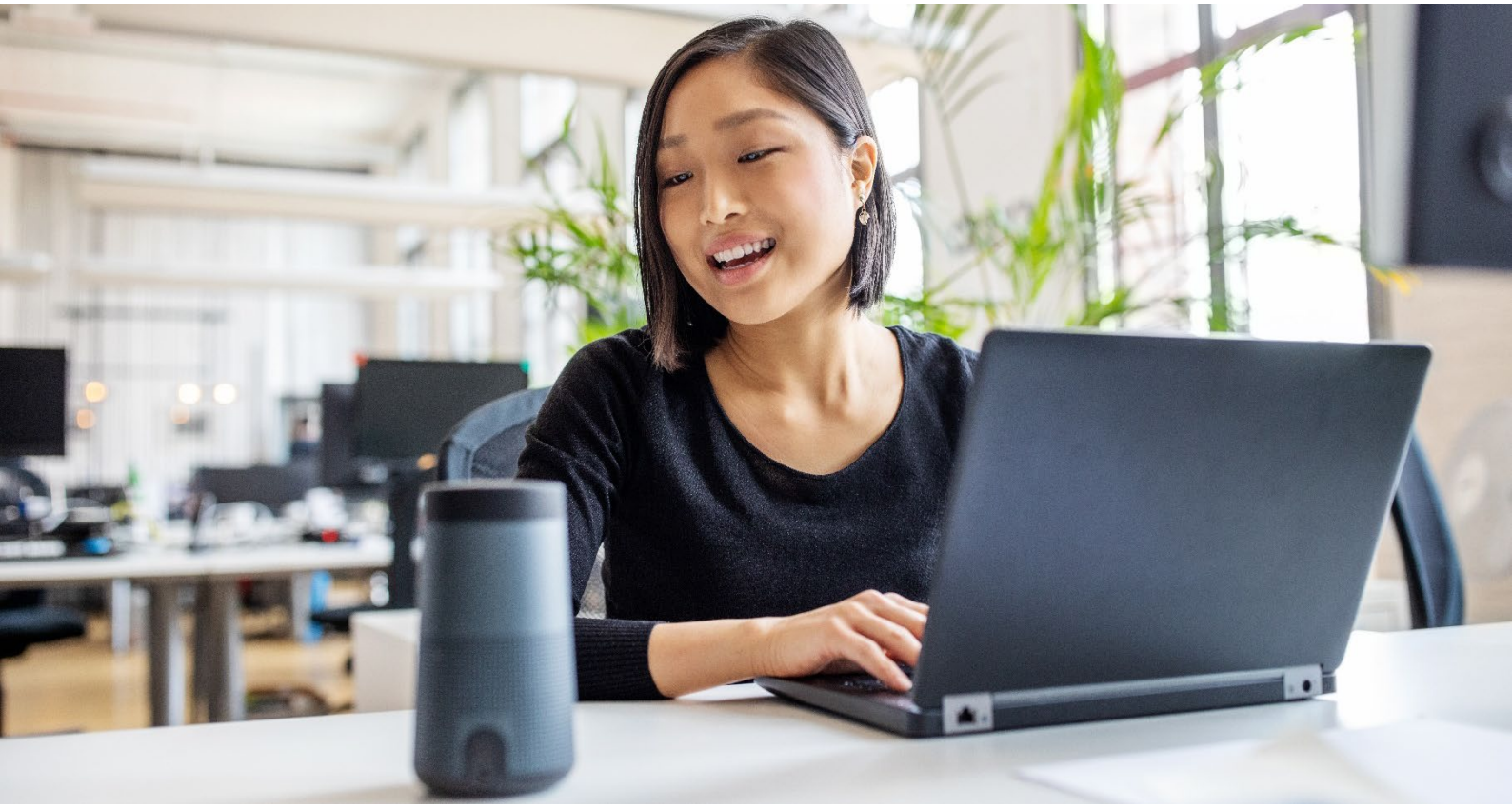
4. 责任主体

《征求意见稿》第五条规定，生成式人工智能产品服务的“提供者”需要承担内容生产者的责任和个人信息处理者的法定责任。该等“提供者”既包括了产品的研发者、利用者，还包括提供可编程接口等技术支持的技术支持者。

然而，对于内容生产者的责任，目前现行的法律法规并未对生成式人工智能产品的生产者责任做出具体规定。类似的规定可以参考《网络信息内容生态治理规定》第二章和第七章中对于网络信息内容生产者的责任规定，但《网络信息内容生态治理规定》是针对网络发布的信息，并不完全等同于AIGC。此外，内容生产者通常指的是在生成式人工智能产品中发布指令或输入内容的用户。生成式人工智能产品根据用户的指令生成内容或结果。因此，用户本身才是实际的内容生产者角色，而提供者只是提供了生成式人工智能产品的工具和服务平台；将提供者代替用户承担内容生产者的责任可能导致用户滥用人工智能产品的情况。

此外，从个人信息处理者的角度来看，生成式人工智能产品的提供者通常是作为用户的受托人角色，根据用户的委托对个人数据进行处理，其并不能“自主决定处理目的、处理方式”。所以，“提供者”并不符合《个人信息保护法》对于“信息处理者”的定义。且根据《个人信息保护法》第五十九条的规定，受委托处理个人信息的受托人承担的是协助个人信息处理者履行义务的责任。然而，《征求意见稿》将提供者的责任提升至与个人信息处理者等同，相当于加重了提供者对个人信息处理的责任范围。

我们认为，对于生成式人工智能产品服务的提供者需要承担的法定责任，应充分考虑到其在内容生产者和个人信息处理者之间的实际角色和责任，并确保其责任范围明确、合理和可操作。此外，还应该明确用户在生成式人工智能产品中的责任和义务，以促进用户的合理使用和防止滥用。



5. 提供者的责任

《征求意见稿》进一步规定了提供者应当履行的一系列义务，主要为：

i. **算法评估和备案**：我们注意到，《征求意见稿》并没有直接设定算法评估和备案责任主体，而采用了引述《具有舆论属性或社会动员能力的互联网信息服务安全评估规定》和《互联网信息服务算法推荐管理规定》的表达方式[1]，而后两者的责任主体分别为“互联网信息服务提供者”及“算法推荐服务提供者”。这两个概念并不等同于《征求意见稿》项下“提供者”。从文意上解释，“提供者”的范围应大于“互联网信息服务提供者”和“算法推荐服务提供者”。但是，《征求意见稿》是要求遵守算法评估和备案。

ii. **人工标注**：提供者在产品研制过程中需要进行人工标注，即向训练数据集添加元数据的过程，以将一般数据转化为人工智能可以识别的结构化数据。这有助于提高生成式人工智能产品的可解释性和可控制性，从而增强产品的透明度和安全性。

iii. **标识义务**：与人工标注不同，标识义务要求提供者在生成内容的合理位置、区域进行显著标识，以区分生成内容与真实内容，避免引起公众混淆。

iv. **实名认证**：提供者应对用户进行实名认证，以确保用户身份真实可靠。这有助于减少虚假身份和恶意行为的发生，提高网络安全和数据保护水平。

v. **防沉迷设置**：提供者应当采取防沉迷设置，以防止未成年人过度使用生成式人工智能产品，保护其身心健康。

vi. **用户投诉接收处理机制**：提供者建立用户投诉接收处理机制，及时受理用户的投诉并进行处理，保障用户的合法权益。

vii. **服务稳定性要求**：提供者应确保其产品和服务的稳定运行，保障用户正常使用权益。但从实际角度上说，由于提供者在研发及应用AIGC产品时，可能需要分析大量的数据，也将面对用户数量的剧增，该等要求可能存在技术障碍，有可能加重提供者的责任。

viii. **模型优化义务**：提供者应对生成式人工智能模型进行优化，以提高其生成结果的质量和安全性，减少可能的违法违规和侵权行为。特别值得强调的是，《征求意见稿》要求对于不符合要求生成的内容，提供者应在3个月内通过模型优化训练等方式防止再次生成。

ix. **科学引导用户义务**：提供者应当在用户使用生成式人工智能产品时进行科学引导，引导用户正确使用产品，防止滥用和违法使用。

结语

尤瓦尔·赫拉利的《未来简史》在总结人类社会的发展规律时，得出的一个重要的观点即为“科技具有向善性”。

因此，在鼓励和发展本世纪最有潜力也极具争议的AI技术的同时，如何对其做出符合“技术向善”精神的监管也成了各国立法者面临的难题。

作为国内首部针对生成式人工智能产品的专项法规，《征求意见稿》标志着国家对于人工智能领域的重视态度。然而，其中的一些细节仍有待进一步完善，许多问题仍需立法者予以明确。从目前的《征求意见稿》来看，其对生成式人工智能服务的提供者规定了较高的合规义务，但目前对于销售者、终端用户等监管和规范似仍处于较低程度的监管，普华永道也将持续关注后续的正式稿的出台。

[1]广东省深圳市南山区人民法院（2019）粤 0305 民初 14010 号民事判决书，裁判日期：2019 年 12 月 24 日。

[2]第六条“利用生成式人工智能产品向公众提供服务前，应当按照《具有舆论属性或社会动员能力的互联网信息服务安全评估规定》向国家网信部门申报安全评估，并按照《互联网信息服务算法推荐管理规定》履行算法备案和变更、注销备案手续。”

联系我们

如您希望就草案或合规建议有更深入的了解，请与我们联系。

普华永道中国

周伟然
全球科技、媒体及通信行业
主管合伙人
wilson.wy.chow@cn.pwc.com
+86 (755) 8261 8886

高建斌
中国内地科技、媒体及通信行业
主管合伙人
gao.jianbin@cn.pwc.com
+86 (21) 2323 3362

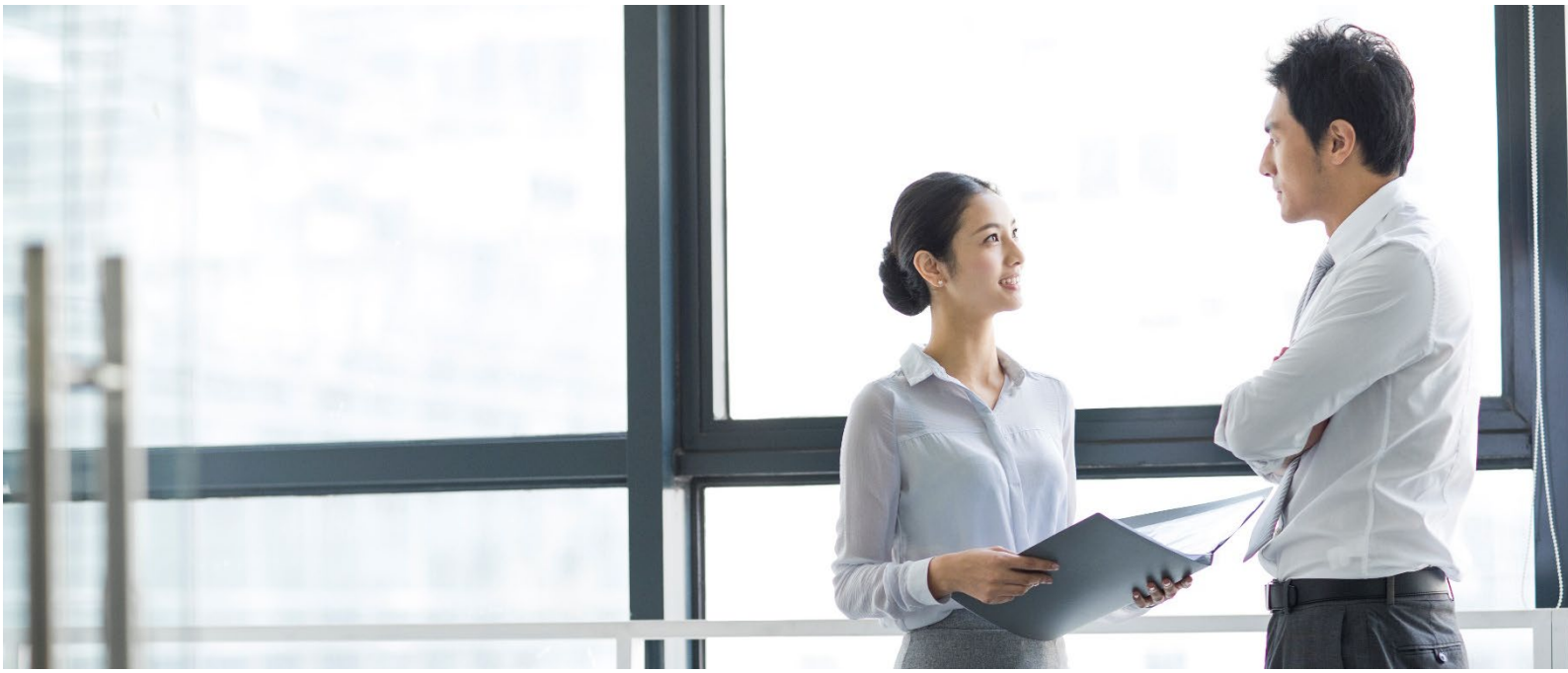
蒋亮
普华永道中国公司及监管服务税务及商务咨询合伙人
liang.l.jiang@cn.pwc.com
+86 (21) 2323 8873

李雨婷
普华永道中国公司及监管服务税务及商务咨询高级顾问
caroline.c.li@cn.pwc.com
+86 (21) 2323 4127

程伟宾律师事务所

董瀚思
程伟宾律师事务所合伙人*
joyce.hs.tung@tiangandpartners.com
+852 2833 4983

李江陵
程伟宾律师事务所合伙人*
chiang.ling.li@tiangandpartners.com
+852 2833 4938



*程伟宾律师事务所是一家独立的香港律师事务所，与普华永道全球网络合作密切。