

促进新加坡金融领域人工智能和数据分析应用中 公平、伦理、问责与透明的原则

新加坡金融管理局 (Monetary Authority of Singapore, MAS)

2018年 (2022年更新)

中文翻译版 · 仅供参考，以英文原文为准

原文来源: <https://www.mas.gov.sg/>

目录

- 引言
- 如何使用本文件
- 范围和定义
- 原则概要
- 公平性
- 伦理
- 问责
- 透明度
- 致谢

1 引言

1.1 本文件包含一套关于在金融产品和服务提供中使用人工智能和数据分析（Artificial Intelligence and Data Analytics，简称"AIDA"）进行决策的通用原则。与人类决策相比，AIDA的性质及其日益广泛的使用可能加剧系统性滥用的风险，导致影响更为广泛、传播速度更快。如果负责任且有效地使用，AIDA在改善业务流程、降低风险和促进更好决策方面具有巨大潜力。

1.2 因此，这套原则的目标如下：

- 为使用AIDA提供金融产品和服务的企业提供一套基本原则供其考量
- 协助企业在自身商业模式和结构中将AIDA治理具体化和可操作化
- 增进公众对AIDA使用的信心和信任

1.3 这套原则由金融行业和其他相关利益方共同创建。在制定这套原则的过程中，MAS通过公平、伦理、问责与透明（Fairness, Ethics, Accountability and Transparency, FEAT）委员会与主要行业利益方紧密合作。MAS还征求了众多金融机构和公司的反馈意见，他们提供了多元的视角和意见。请参阅第9节的致谢。

1.4 在制定这套原则的过程中，MAS与个人数据保护委员会（Personal Data Protection Commission, PDPC）和资讯通信媒体发展局（Infocomm Media Development Authority, IMDA）紧密合作，使本原则与IMDA的AI治理举措保持一致，包括人工智能和数据伦理使用咨询委员会（Advisory Council on the Ethical Use of AI and Data）以及总体AI治理框架（AI Governance Framework）。这些举措旨在帮助企业以负责任的方式开发和部署AI技术。PDPC的《AI与个人数据讨论文件》（Discussion Paper on AI and Personal Data）是本文件的补充，并就选择适当的组织治理措施和采用良好的数据管理实践等领域提供了额外的说明和指导。

2 如何使用本文件

2.1 本文件为企业在评估现有或制定新的内部框架以治理AIDA使用时提供了应考量的原则。这套原则不旨在取代现有的相关内部治理框架。企业应继续遵守所有适用的法律和要求。

2.2 这套原则不旨在具有规定性。本文件中提供的示例仅为说明性质。尽管如此，通过持续的行业参与过程，MAS有可能确定更具体或更具技术性的指导将有利于行业的领域，而这套原则可作为考量和讨论的基础框架。

2.3 在本文件所列的所有原则中，企业可以根据AIDA驱动决策的重要性（materiality）来校准其内部治理框架下的行动和要求。以下提供了在确定重要性时可使用的非详尽的示例性考量因素：

- AIDA在决策中的使用程度；
- AIDA模型的复杂性；
- AIDA驱动决策过程的自动化程度；
- 对不同利益相关者（包括个人）的影响的严重性和可能性；
- 金融和财务影响；
- 监管影响；
- 可用的救济选项。

2.4 由于各原则相互关联和强化，建议读者从整体上看待这些原则。

3 范围和定义

3.1 这套原则可指导所有使用AIDA提供金融产品和服务的企业。

3.2 "AIDA" 是指人工智能或数据分析（Artificial Intelligence or Data Analytics），定义为协助或替代人类决策的技术。虽然这一定义较为宽泛，但本文件中的原则可适用于复杂和简单的AIDA技术。使用AIDA的企业应根据第2.3节所述的重要性考量来校准其内部治理框架。

3.3 本文件使用以下定义：

- "AIDA" 是指人工智能或数据分析，定义为协助或替代人类决策的技术。
- "AIDA企业"（AIDA Firms）是指使用AIDA提供金融产品和服务的企业。
- "AIDA驱动的决策"（AIDA-driven decisions）是指由AIDA做出或支持的决策及决策结果。
- "AIDA驱动的决策过程"（AIDA-driven decision-making）是指使用AIDA进行决策的过程。
- "数据主体"（Data subjects）是指受到或可能受到AIDA驱动的决策影响的个人和/或企业。
- "模型"（Models）是指用于得出AIDA驱动的决策的特定AIDA算法和算法过程。

4 原则概要

公平性 (Fairness)

合理性 (Justifiability)

1. 个人或群体不会通过AIDA驱动的决策被系统性地置于不利地位，除非这些决策具有合理性。
2. 将个人属性作为AIDA驱动决策的输入因素须有合理依据。

准确性与偏差 (Accuracy and Bias)

1. 用于AIDA驱动决策的数据和模型定期接受准确性和相关性审查与验证，并最大限度地减少无意偏差。
2. AIDA驱动的决策定期接受审查，以确保模型按设计和预期运行。

伦理 (Ethics)

1. AIDA的使用与企业的伦理标准、价值观和行为准则保持一致。
2. AIDA驱动的决策至少应达到与人类驱动决策相同的伦理标准。

问责 (Accountability)

内部问责 (Internal Accountability)

1. 在AIDA驱动的决策过程中使用AIDA须经适当的内部权限机构批准。
2. 使用AIDA的企业对内部开发和外部采购的AIDA模型均须承担责任。
3. 使用AIDA的企业主动提高管理层和董事会对其AIDA使用的认知。

外部问责 (External Accountability)

1. 为数据主体提供渠道，使其能够就影响其自身的AIDA驱动决策进行咨询、提交申诉和请求审查。
2. 在对AIDA驱动的决策进行审查时，应考虑数据主体提供的经核实且相关的补充数据。

透明度 (Transparency)

1. 为增强公众信心，作为日常沟通的一部分，主动向数据主体披露AIDA的使用情况。
2. 应数据主体的请求，提供关于使用哪些数据对其做出AIDA驱动的决策以及数据如何影响决策的清晰解释。
3. 应数据主体的请求，提供关于AIDA驱动的决策可能对其产生的后果的清晰解释。

5 公平性

5.1 AIDA驱动的决策不在没有合理理由的情况下使任何特定个人或群体处于不利地位，这一点至关重要。如今，对不同群体提供差异化待遇是一种现有做法，也是当前商业模式的一部分。虽然AIDA的使用可以实现基于数据分割和聚类的分析，但这也意味着群体间的差异化可能以更大的规模和更快的速度发生。AIDA的使用还可能创造出识别或分析以前无法进行的新颖差异化的能力。如果管理不当，这可能在系统层面固化不合理的差异化。

5.2 AIDA企业可能已建立内部治理框架，用于评估和设定AIDA驱动决策（包括使用特定数据属性）背后适当的合理性标准。此类内部治理框架也可能具有评估用于AIDA驱动决策的数据和模型准确性的流程，并可根据决策的重要性进行校准。

5.3 公平性原则关注两个关键方面——首先是合理性，其次是AIDA驱动决策的准确性和偏差。

合理性

1. 个人或群体不会通过AIDA驱动的决策被系统性地置于不利地位，除非这些决策具有合理性。
2. 将个人属性作为AIDA驱动决策的输入因素须有合理依据。

5.4 **案例说明1：** 一家AIDA企业使用AIDA根据多种因素决定是否批准客户增加信用卡额度。该企业根据其内部治理框架，向内部审批机构提供使用每个因素的理由说明。该企业还说明每个因素或不同权重的因素组合可能如何影响决策。基于这些理由说明，一些客户可能获得信用卡额度增加的批准，而另一些则可能不会。

5.5 案例说明2： 一家AIDA企业使用AIDA进行客户细分，以便为不同客户群体提供更相关的产品服务。这可能包括使用客户的年龄（一个个人属性）作为输入因素，以决定是否提供与退休相关的金融服务和产品。根据该AIDA企业的内部治理框架，鉴于年龄与退休之间的关联，使用年龄作为该AIDA驱动决策的输入因素可能是合理的。

准确性与偏差

3. 用于AIDA驱动决策的数据和模型定期接受准确性和相关性审查与验证，并最大限度地减少无意偏差。

4. AIDA驱动的决策定期接受审查，以确保模型按设计和预期运行。

5.6 案例说明3： 根据内部治理框架，一家AIDA企业制定了模型验证演练和错误或无意偏差测试的要求。演练或测试的频率取决于决策的重要性和模型的复杂性。对于特别重要和复杂的深度学习模型，AIDA企业要求更频繁的验证和测试。相反，对于重要性较低和较简单的模型，每季度、每年甚至更低频率的验证和测试即可。

5.7 案例说明4： 除定期审查外，AIDA企业的内部治理框架还确定了特定的审查触发条件，包括使基本模型假设无效的意外变化。例如，关于某实体是否对其债务违约的意外法院判决可能触发对信用风险参数的审查，以确定是否需要修订参数。

6 伦理

6.1 所有企业，无论是否使用AIDA，都按照其伦理标准运营。这些伦理标准通过多种方式表达，包括公司价值观、行为准则和使命宣言，且可能因企业和地区而异。遵守这些伦理标准同样适用于AIDA的使用。

5. AIDA的使用与企业的伦理标准、价值观和行为准则保持一致。

6. AIDA驱动的决策至少应达到与人类驱动决策相同的伦理标准。

6.2 案例说明5： 一家AIDA企业的人力资源部门有现行的招聘规则，规定性别不应作为招聘决策的考量因素，这与其关于多元化和公平就业的整体公司政策一致。该AIDA企业决定实施一个模型

或算法来协助筛选和遴选潜在候选人。根据该AIDA企业的内部治理框架，为保持与人类驱动决策相同的伦理标准，该模型或算法不使用性别作为输入或派生属性。

6.3 案例说明6： 一家AIDA企业没有明确的机构来监督和维护其伦理标准，因此决定设立一个全组织范围的伦理审查委员会。此外，该AIDA企业还考虑设立针对特定项目的跨职能伦理审查委员会，由具有相关专业知识的利益相关者和独立代表组成，向全组织范围的伦理审查委员会汇报。

7 问责

7.1 在AIDA企业内部，对AIDA驱动的决策有明确的责任和所有权，并由适当的内部审批机构批准AIDA的使用，这一点至关重要。此类问责适用于AIDA的所有使用，无论是内部开发还是外部采购。

7.2 AIDA驱动的决策，特别是那些可能对数据主体产生重大影响的决策，需要基于对数据主体尽可能准确的理解。数据主体可以通过旨在促进沟通和寻求救济的渠道提供关于自身的准确信息。

7.3 问责原则因此关注两个关键方面——首先是内部问责，涉及AIDA企业的内部治理；其次是外部问责，涉及AIDA企业对数据主体的责任。

内部问责

7. 在AIDA驱动的决策过程中使用AIDA须经适当的内部权限机构批准。

8. 使用AIDA的企业对内部开发和外部采购的AIDA模型均须承担责任。

9. 使用AIDA的企业主动提高管理层和董事会对其AIDA使用的认知。

7.4 案例说明7： 一家企业目前已有决策的内部审批结构，在设定AIDA驱动决策的审批权限时选择依赖该结构。例如，目前金融市场主管批准与交易职能相关的所有决策。如果该职能部门希望引入用于执行交易的AIDA模型，金融市场主管将根据AIDA企业的内部治理框架同样批准该模型的使用。

7.5 案例说明8： 在设定AIDA驱动决策的审批权限时，企业的内部治理框架考虑了决策的重要性和复杂性。对于高度重要的决策，企业将审批级别设在首席执行官或董事会层面，并进行适当的尽职调查，以确保审批机构对用于决策的数据和模型逻辑有充分的了解。

7.6 案例说明9： 一家AIDA企业定期向董事会更新企业内部AIDA的使用情况，以便董事会对所有重要的AIDA驱动决策保持全面的了解。作为良好实践，该企业组织定期的员工大会以提高对企业内AIDA使用的认知，其中可能包括部门案例研究或学习要点分享。

外部问责

10. 为数据主体提供渠道，使其能够就影响其自身的AIDA驱动决策进行咨询、提交申诉和请求审查。

11. 在对AIDA驱动的决策进行审查时，应考虑数据主体提供的经核实且相关的补充数据。

7.7 案例说明10： 在通知数据主体AIDA驱动的决策时，AIDA企业提供联系电话和电子邮件地址，以便数据主体咨询该决策并提供合理的审查依据。该企业调整现有的客户咨询处理框架，以评估问题的重要性和回应此类请求的可行性。

7.8 案例说明11： AIDA企业为数据主体提供在线数据管理工具，如隐私或个人数据仪表盘。通过这些仪表盘，数据主体可以选择审查、更新或编辑其信息，以确保准确性和相关性。

8 透明度

8.1 虽然提高AIDA企业在AIDA使用方面的透明度可以增进公众的理解和信心，但过度的透明可能造成混淆，或为个人利用或操纵AIDA模型创造意外机会。在确定AIDA使用的适当透明度水平时，平衡这些考量因素非常重要。在确定透明度水平时，决策的重要性也是相关因素。

8.2 数据主体可能寻求关于用于决策的数据、数据如何影响决策以及对数据主体的后果的清晰解释。清晰的解释不需要暴露知识产权或公开专有源代码。清晰的解释可以侧重于促进数据主体对企业AIDA使用的理解。

12. 为增强公众信心，作为日常沟通的一部分，主动向数据主体披露AIDA的使用情况。

13. 应数据主体的请求，提供关于使用哪些数据对其做出AIDA驱动的决策以及数据如何影响决策的清晰解释。

14. 应数据主体的请求，提供关于AIDA驱动的决策可能对其产生的后果的清晰解释。

8.3 案例说明12： 一家提供汽车保险的AIDA企业告知客户将使用驾驶模式数据来审查客户保费。该企业解释某些驾驶模式如何影响保费，提供说明性图表来解释识别的模式及其对不同客户群体的影响。该企业还开发了用户界面，允许客户调整潜在输入以了解对保费结果的影响。被收取较高保费的客户可以联系该企业获得关于保费增加原因的有意义的解释。AIDA企业与客户分享其所属的客户群体及对保费的相应影响。AIDA企业还广泛地与客户分享用于分析其驾驶模式的数据，并为客户提供提交更准确或更新信息的途径。从长远来看，客户可以选择改变其行为并养成良好的驾驶习惯，以确保保费不会增加。

8.4 案例说明13： 一家AIDA企业在确定适当的透明度水平时考虑AIDA驱动决策的重要性。例如，该AIDA企业使用AIDA模型进行欺诈检测和识别可能的"红旗"。鉴于重要性考量和对模型被操纵或利用的担忧，AIDA企业决定不分享关于所使用的AIDA模型或提供与此领域相关的解释。另一方面，该AIDA企业部署AI电话接线员处理客户来电并直接与客户互动。在这种情况下，AIDA企业选择在对话开始时告知数据主体他们正在与AI电话接线员互动。

9 致谢

MAS感谢FEAT委员会在FEAT原则的创建和发展中发挥的关键作用。

- **谢福华** (Hsieh Fu Hua, 联合主席)，PrimePartners联合创始人兼顾问
- **V K Rajah** (特别顾问)，资深大律师，Essex Court Chambers Duxton (新加坡集团) 成员
- **张秀莲** (Teo Swee Lian, 顾问)，新电信 (SingTel) 非执行独立董事，Avanda Investment Management非执行独立董事
- **Raymond Au**，安联 (Allianz) 首席数据科学家兼亚洲实验室主管
- **Paul Cobban**，星展银行 (DBS) 首席数据与转型官
- **Shameek Kundu**，渣打银行 (Standard Chartered Bank) 首席数据官
- **Richard Lowe**，大华银行 (UOB) 首席数据官
- **Donald MacDonald**，华侨银行 (OCBC Bank) 集团客户分析与决策主管

MAS感谢以下贡献者为FEAT原则提供的宝贵意见和反馈 (按英文字母排序)：

Amazon Web Services、American Express Int'l Inc、亚洲云计算协会（Asia Cloud Computing Association）、英杰华（Aviva Ltd）、巴克莱银行（Barclays Bank PLC）、CFA Institute、CFA Society Singapore、花旗银行（Citibank NA）、瑞信（Credit Suisse AG）、德意志银行新加坡分行（Deutsche Bank AG Singapore）、安盛天平保险（Etiqa Insurance Pte Ltd）、大东方普通保险（Great Eastern General Insurance Ltd）、大东方人寿（Great Eastern Life Assurance Co Ltd）、汇丰银行（HSBC）、新加坡投资管理协会（Investment Management Association of Singapore）、Kristal Advisors、新加坡人寿保险协会（Life Insurance Association, Singapore）、劳合社亚洲（Lloyd's of London (Asia) Pte Ltd）、陆金所国际（Lu International）、万事达卡（Mastercard）、微软（Microsoft）、新加坡电子支付网络（Network for Electronic Payments (Singapore) Pte Ltd）、奥纬咨询（Oliver Wyman）、华侨银行（OCBC Bank）、保诚新加坡（Prudential Assurance Company Singapore (Pte) Limited）、丰业银行（Scotiabank）、新加坡证券业协会（Securities Association of Singapore）、新加坡交易所（Singapore Exchange）、新加坡金融科技协会（Singapore FinTech Association）、渣打银行（Standard Chartered Bank）、新加坡银行公会（The Association of Banks in Singapore）、加拿大皇后大学史蒂芬·J·R·史密斯商学院（The Stephen J.R. Smith School of Business at Canada's Queen's University）、东京海上人寿保险新加坡（Tokio Marine Life Insurance Singapore Ltd）、大华银行（United Overseas Bank Limited）、Yair Zick（新加坡国立大学）。