



EPOS |

每个人，各有来处。

EPOS的解释对象

EPOS解释的对象不是一个脱离情境的人格标签，而是具体情境中的行动选择。一个完整的解释需要说明，当事人在当时注意到了什么，如何理解这些信息，看见了哪些可行选项，各个选项分别意味着什么，以及某个选项为什么获得了最高主观优先级。

性格分类能够描述长期倾向，却不能单独说明一次行为为何在此时此地发生。谨慎、探索、完美主义、控制需求等词语可以概括现象，但只有在它们与具体事件、候选行动和结果预期建立联系后，才成为解释的一部分。

EPOS将动机定义为使某个行动在当时获得更高主观优先级的价值关系。动机可以表现为获得某种结果、避免某种损失、保护某段关系、维持控制、确认未知，或尽快结束难以承受的状态。来源则是使这种价值排序更容易出现的形成条件。

解释的最低结构由事实、竞争解释、关键取舍、形成来源、翻转条件和未来预测构成。反事实检验是其中的必要部分。若一个解释不能说明关键条件改变后行为会怎样变化，它仍停留在描述层面。反事实成立也不等于因果关系已经得到证明，还需要新的行为记录继续检验。

行为的四层来源

EPOS把行为来源分为稳定倾向、学习历史、当下状态和事件结构。四层来源不是互相排斥的原因，而是共同决定信息如何被注意、意义如何形成、行动如何排序的条件。

稳定倾向

稳定倾向指较长时间内相对一致的刺激敏感度、奖赏偏好、风险反应、控制需要、社交取向和行动节奏。它可能受到遗传、发育和长期环境的共同影响，但不能由单个生物指标、身体特征或一份问卷直接测出。稳定倾向改变的是某些线索更容易被注意、某些结果更容易被赋予高权重的概率，而不是预先决定具体行动。

学习历史

学习历史包括过去的成功、失败、奖惩、失控、关系经验、规则和习惯。经历留下的不是简单结论，而是对线索、行动和后果之间关系的更新。一次强烈的成功、羞耻或失控，可能比多次平淡经历留下更高权重。相似经历不会必然产生相同结果，因为经历发生时的年龄、解释方式、支持系统和后续反馈并不相同。

当下状态

疲劳、压力、时间紧迫、情绪唤醒、身体不适和可用资源会暂时改变注意范围、耐心和风险判断。短期状态可以使同一个人在相似情境中做出不同选择。高压可能缩短搜索过程，使熟悉反应更快胜出；资源充足时，更多候选行动才有机会进入比较。

事件结构

事件结构包括可见选项、信息完整度、后果是否可逆、行动窗口、谁在场、社会规则 and 不同价值之间的冲突。情境不是背景，而是行为生成过程的一部分。无法撤回的提交、突然改变的安排、公开评价和私下交流，即使涉及同一件事，也会形成不同的行动结构。

行动选择机制

行为生成从选择性注意开始。事件中只有一部分信息进入后续处理。截止时间、语气、机会、损失、关系变化或身体感受中的某一部分先被捕捉，未被注意的信息难以影响当下行动。

进入注意的信息随后获得意义。同一句通知可以被理解为普通调整、失去控制、被强迫、一次必须避免的损失，或关系中地位的变化。意义并非由刺激独立决定，而是由四层来源共同塑造。

意义形成后，当事人生成主观可见的行动集合。继续、停止、求助、退出、反击、拖延、隐藏和重新确认都可能成为候选行动。客观可行的办法不一定出现在主观行动集合中，尤其在压力升高、时间不足或经验有限时。

每个候选行动都会同时带来收益和代价。EPOS优先检查六类价值信号：获得与新奇、威胁与损失、控制与可预测、关系与归属、地位与竞争、体力时间与认知成本。这六类信号不是六种人格，而是可以同时进入一次选择的六类价值。

行动选择可以用下式表达。A代表当事人当时能够看见的行动集合；T、L、S、E分别代表稳定倾向、学习历史、当下状态和事件结构；u表示某个行动在特定价值维度上的主观结果；w表示该价值此刻的权重；C表示行动成本；ε表示尚未观测的影响。

$$a^* = \operatorname{argmax}_{a \in A} [\sum w(k, T, L, S, E) \times u(k, a, E) - C(a, E) + \epsilon]$$

该表达式用于明确模型结构，不是已经完成参数估计的个人评分公式。当某个选项明显领先时，行动更容易发生；当两个互不相容的选项都具有高价值、差距又很小时，犹豫、僵住或退出可能成为新的行动结果。行动产生的结果会再次写入学习历史，改变下一次选择的权重。

动机识别方法

EPOS不从行为直接跳到单一结论。相同行为可能由不同动机产生，单一动机也可能通过不同动作表现。因此，解释过程先建立多个能够说明当前事实、但对关键条件具有不同预测的竞争解释。

解释开始时先区分事件事实、当事人的即时体验和事后理解。事实包括发生了什么、采取了什么行动、当时还能做什么、已知后果是什么。即时体验包括最先出现的注意、感受、身体变化和行动冲动。事后理解可以提供来源线索，但不能替代当时事实。

候选解释通常保持在两个至四个。每个解释都必须写出支持证据、反证和未知项。无法产生不同预测的两个解释应合并；只是在重新命名行为的解释应删除。

后续信息的选择以区分度为中心。优先获取能够使至少两种解释得到不同支持的信息，同时控制回答负担。每一轮新信息都应增加、降低或维持某个解释的支持，而不是只扩充故事长度。

停止条件不是固定问满全部问题，而是事实链、主要取舍、关键来源和翻转条件已经足以形成可检验解释。信息不足时保留竞争解释，不使用更流畅的语言掩盖不确定性。

解释的组成

一次完整解释首先给出当前最有支持的直接原因，随后还原从注意、意义、候选行动到价值排序的过程。核心动机应被写成具体取舍，即行动试图获得、保护、避免或结束的状态，而不是写成抽象人格名称。

稳定倾向与重复模式只有在存在跨事件证据时进入解释。形成来源由学习历史、长期环境和可能的先天倾向共同构成。任何来源判断都需要标明它来自当事人的明确叙述、重复行为，还是仍待检验的假设。

翻转条件描述哪项变化最可能使行动排序改变。未来预测由翻转条件推导，针对相似而尚未发生的情境。预测必须写在结果出现之前，避免用事后解释替代预测。

解释末端保留仍不能确定的内容，包括证据不足的来源、仍有支持的竞争解释、未观察到的选项和能够继续检验的条件。这一结构使解释能够随新事实更新，而不是成为不可修改的人格判决。

证据等级

事实层由当事人直接叙述、行为记录或多次重复出现的模式构成。事实可以明确陈述，但仍需区分当时记录与事后回忆。

推断层由能够同时解释多项事实、并排除部分替代解释的判断构成。推断需要指认依据并保留置信程度。

假设层由可能解释现象但个人证据不足的机制构成。假设必须保留为可验证内容，不能改写成已经发生的事实。神经机制、生物差异和成长来源通常属于这一层，除非存在直接而充分的个体证据。

证据等级可以随新材料上升或下降。一次重复不自动使假设成为事实，多次相似叙述也可能来自同一种记忆偏差。跨情境重复、独立记录和预先预测的行为比事后认同提供更强证据。

问卷与九种角色

问卷描述一个人在多类情境中的默认策略，不直接解释单次行为。EPOS保留九种功能角色，用于压缩和呈现策略差异。角色分别对应探路、抢先行动、防止出事、把事做好、提前规划、组织行动、维持合作、照护系统和传授经验。它们代表九种解决问题的常见功能，不代表九种高低不同的人。

主角色表示当前题库和评分规则下最稳定的功能接近度，翼型表示与主角色相邻且次高的策略。角色边界不是自然界已经得到证明的离散类别。任何角色都不能单独推出具体情境中的行为，也不能推出激素水平、脑结构、医学状态或个人能力上限。

题目使用具体生活场景，使被测者比较自己在机会、风险、信息不完整、冲突和资源限制中的实际反应。题目需要同时控制社会赞许、语言难度、重复含义和情境偏差。40题与90题版本的差别主要是覆盖范围与稳定性，不是两套不同理论。

角色结构需要通过内部一致性、重测稳定性、题目区分度、角色边界和外部关联进行检验。与大五等成熟维度应存在可解释的关联，同时不能完全等同，否则九种角色只是在重新包装已有分类。

长期来源建模

单次事件能够说明一次行动的形成过程，但难以可靠确定一个人长期反复出现的来源。长期来源建模通过跨会话累积行为模式、情绪线索、价值取舍、关键经历和自我认知，观察某种排序是否在不同情境中重复出现。

每条长期记录包含类型、时间、情境、原始表述、来源会话和置信程度。单次陈述只提供有限支持；跨时间、跨情境重复出现的模式获得更高权重。自我描述与实际事件不一致时，两者都应保留，不能自动选择更符合既有画像的一方。

长期画像不是固定总结。新事件可以强化旧判断，也可以使旧判断降级或失效。画像更新必须保留证据链和版本变化，使任何结论都能追溯到具体材料。

个人长期记录与共享案例库相互分离。个人记录只服务于本人；进入共享研究、检索和评测的案例需要单独同意、脱敏和审查。

人工智能系统结构

EPOS人工智能系统由基础语言模型、解释协议、个人记忆、案例检索、规则版本和评测集组成。基础语言模型负责理解自然语言、生成候选解释和组织表达，不自动拥有EPOS的理论边界。

解释协议限定事实提取、竞争解释、信息选择、反事实检验、证据等级和输出结构。协议的作用是减少过早结论、空泛标签、未经证据的脑区故事和固定人格决定论。

案例库保存经同意、脱敏和标注的事件、解释过程、修正意见和结果反馈。新案例不会自动改写基础模型参数，也不会直接影响下一次回答。未经审查的实时学习会把误解、恶意输入、隐私和偏见一并放大。

案例影响模型的路径依次为收集、脱敏、审查、标注、加入检索或评测、修改规则、与旧版本比较和发布。样本规模、标签质量和资源满足要求后，才评估微调或本地模型。

生物层的使用边界

EPOS最初关注神经回路、激素水平和大脑结构差异对行为的影响。现代证据支持生物因素参与个体差异，但不支持从单一指标确定个人性格或行为。生物层在EPOS中作为稳定倾向和状态变化的可能来源，不作为问卷结果的直接输出。

群体层面的脑与行为关联通常较小，可重复研究往往需要大样本。不能依据单个脑区大小、激活强度或结构差异对个人行为下结论。

激素作用依赖情境、测量时间和个体差异。以睾酮为例，它与攻击行为的总体关联较小，不能简化为高睾酮必然带来攻击或冒险。

指长比不能可靠测量个人当前睾酮水平，不进入问卷、角色计算和个人解释。上行网状激活系统与内外向的早期理论保留为性格生物学历史材料，不作为现代个体解释的单一机制。

古老社会分工可以提供探索、防御、组织、照护和传授等互补功能的叙事来源，也可以形成研究假设。现有证据不足以证明古老分工直接形成现代九种角色，因此它不进入个人因果结论。

验证方法

EPOS的有效性不能只由当事人觉得准确来判断。普遍适用、表达流畅的句子也可能产生强烈认同。验证需要同时检查信息效率、证据忠实度、解释区分度、翻转条件、未来预测和置信校准。

信息效率记录形成一个可用解释所需的信息量，以及无效和重复信息的比例。证据忠实度检查解释中的事实是否真实存在于原始材料，是否混入模型生成但当事人未提供的信息。

解释区分度通过材料打乱、报告盲选和通用表述基线检验。若不同人的报告可以轻易互换，模型只是生成了普适描述。

翻转条件在关键条件变化时检验行为是否按预测改变。未来预测需要在结果出现前保存，并与简单人格标签、通用语言模型和朴素基线比较。解释正确率、预测准确率和置信程度分别记录，不能互相替代。

问卷验证包括题目一致性、两至四周重测稳定性、角色边界、与成熟人格维度的重合和差异。长期来源建模还需检查引用证据的真实性、画像更新方向和跨时间预测增量。

失败案例需要按类型保存，包括事实还原错误、候选解释缺失、信息选择无效、错误来源推断、翻转条件失效、置信过高和表达造成误解。规则更新以减少已知失败为目标，同时检查是否引入新的失败。

隐私与使用边界

行为解释涉及个人经历、关系、情绪和长期模式。提供解释服务与案例研究需要分开授权；拒绝案例研究不影响基本解释。默认不收集真实姓名、精确住址、身份证件和无必要的联系方式。

案例进入共享库前删除人名、学校、公司、精确地点及可组合识别信息。个人长期记录与共享案例库分别存储并实施访问控制。数据保存期限、删除方式、访问日志、速率限制和密钥隔离属于系统的基本组成。

EPOS用于自我理解与一般行为反思，不用于心理疾病诊断、药物建议、法律意见和危机干预。它不得用于招聘、保险、授信、司法、教育机会分配，或在本人不知情时分析第三方。角色、生物指标、性别和群体倾向不能成为降低个人机会的理由。

结语

EPOS把行为解释建立在具体事件、价值排序和可检验条件之上。稳定倾向、学习历史、当下状态和事件结构共同决定哪些信息进入注意、哪些意义首先出现、哪些行动能够被看见，以及哪个选项最终胜出。

模型不把分类名称、脑区、激素或单段成长经历当作终点。解释需要比较竞争原因，说明关键取舍，保留证据等级，并对条件变化和未来行为作出能够失败的预测。

每个结论有来源，每个推断有等级，每个假设允许被推翻。EPOS对行为的理解由此成为一套能够积累、修正和检验的结构。

参考资料

- [1] Mischel, W. 与 Shoda, Y. (1995) 。《人格的认知-情感系统理论》。Psychological Review, 102(2), 246-268。doi:10.1037/0033-295X.102.2.246
- [2] Fleeson, W. (2001) 。《人格特质作为状态分布》。Journal of Personality and Social Psychology, 80(6), 1011-1027。doi:10.1037/0022-3514.80.6.1011
- [3] Rangel, A., Camerer, C. 与 Montague, P. R. (2008) 。《价值决策神经生物学研究框架》。Nature Reviews Neuroscience, 9, 545-556。doi:10.1038/nrn2357
- [4] Cisek, P. (2007) 。《行动选择的皮层机制：可供性竞争假说》。Philosophical Transactions of the Royal Society B, 362, 1585-1599。doi:10.1098/rstb.2007.2054
- [5] Dolan, R. J. 与 Dayan, P. (2013) 。《大脑中的目标与习惯》。Neuron, 80(2), 312-325。doi:10.1016/j.neuron.2013.09.007
- [6] Wood, W. 与 R  nger, D. (2016) 。《习惯心理学》。Annual Review of Psychology, 67, 289-314。doi:10.1146/annurev-psych-122414-033417
- [7] Schwabe, L. 与 Wolf, O. T. (2009) 。《压力促使人类采用习惯性行为》。The Journal of Neuroscience, 29(22), 7191-7198。doi:10.1523/JNEUROSCI.0979-09.2009
- [8] Lerner, J. S. 等 (2015) 。《情绪与决策》。Annual Review of Psychology, 66, 799-823。doi:10.1146/annurev-psych-010213-115043
- [9] Polderman, T. J. C. 等 (2015) 。《五十年双生子研究中的人类特质遗传率元分析》。Nature Genetics, 47, 702-709。doi:10.1038/ng.3285
- [10] Marek, S. 等 (2022) 。《可重复的全脑关联研究需要数千名参与者》。Nature, 603, 654-660。doi:10.1038/s41586-022-04492-9

- [11] Eisenegger, C., Haushofer, J. 与 Fehr, E. (2011) 。《睾酮在社会互动中的作用》。Trends in Cognitive Sciences, 15(6), 263-271。doi:10.1016/j.tics.2011.04.008
- [12] Geniole, S. N. 等 (2020) 。《睾酮是否与人类攻击行为相关：元分析检验》。Hormones and Behavior, 123, 104644。doi:10.1016/j.yhbeh.2019.104644
- [13] Hönekopp, J. 等 (2007) 。《指长比与成人性激素水平：新数据与元分析综述》。Psychoneuroendocrinology, 32(4), 313-321。doi:10.1016/j.psyneuen.2007.01.007
- [14] Yarkoni, T. 与 Westfall, J. (2017) 。《心理学应选择预测而非解释：来自机器学习的启示》。Perspectives on Psychological Science, 12(6), 1100-1122。doi:10.1177/1745691617693393
- [15] Miller, T. (2019) 。《人工智能中的解释：来自社会科学的启示》。Artificial Intelligence, 267, 1-38。doi:10.1016/j.artint.2018.07.007
- [16] Rozenblit, L. 与 Keil, F. (2002) 。《被误解的常识科学边界：解释深度错觉》。Cognitive Science, 26(5), 521-562。doi:10.1207/s15516709cog2605_1