arXiv:2601.02407v1 [cs.NE] 1 Jan 2026

混沌中的人格演化：一种用于环境压力下迭代囚徒困境角色发现的大语言模型增强框架

Oğuzhan YILDIRIM伊兹密尔理
工学院计算机工程系
oyildirim@std.iyte.edu.tr2025
年12月

摘要

迭代囚徒困境（IPD）的标准模拟是在确定性、无噪声环境中运行的，由此产生的策略或许在理论上是最优的，但在面对现实世界的不确定性时却十分脆弱。本文探讨演化博弈论研究中的两个关键缺口：（1）策略演化过程中缺乏真实的环境压力因素；（2）存在“可解释性鸿沟”，进化而来的遗传策略仍是缺乏语义含义的不透明二进制序列。我们提出一种新颖框架，将随机环境扰动（“上帝模式”）与基于大语言模型（LLM）的行为画像相结合，把进化基因型转化为可解释的角色原型。实验结果表明，在混沌环境下演化出的策略具有更强的韧性，并呈现出鲜明的行为表型——从“无情资本家”到“外交型执法者”——大语言模型能够轻松识别这些表型，但仅凭人工基因组检查几乎无法解读。本研究将演化计算与可解释人工智能相结合，为多智能体系统中的自动化智能体表征提供了一个范例。

关键词：迭代囚徒困境、遗传算法、大语言模型、可解释人工智能、环境压力因素、博弈论、多智能体系统

1 引言

1.1 背景

迭代囚徒困境（迭代囚徒困境）自阿克塞尔罗德的开创性锦标赛以来，一直是理解合作如何涌现的基础范式 [2]。阿克塞尔罗德证明，像以牙还牙这样简单的策略能够胜过更复杂的方法，并确立了互惠、宽恕和报复等原则，这些原则至今仍为协作式人工智能研究提供参考。林德格伦随后开展的研究 [7] 将遗传算法引入迭代囚徒困境，使策略能够通过选择、交叉和变异实现进化——这一方法论此后已成为演化博弈论中的标准做法。

然而，当代迭代囚徒困境研究仍存在两个显著局限：

局限 1：无菌环境。绝大多数演化迭代囚徒困境模拟都在理想化条件下运行，行动能够完美执行，收益始终保持不变。相比之下，现实世界中的互动存在噪声、不确定性和环境波动。“颤抖之手”修正 [10]承认参与者偶尔会犯错，但很少有演化框架会在策略演化过程中系统地纳入多个同时作用的应激因子。

局限 2：可解释性鸿沟。即使遗传算法成功演化出高性能策略，其产生的基因型仍是不透明的二进制序列（例如，`[1,0,1,1,0,1,0,0,1,1,0,1,1,0,1,0,1,1]`）。研究人员能够观察到某种策略确实有效，却难以用人类能够理解的术语阐明原因。

1.2 我们的贡献

本文提出了一个综合框架，旨在解决这两项局限：

1. **随机环境压力因素（“上帝模式”）：**我们引入了五种概率性扰动——颤抖之手、经济危机、高诱惑、记忆丧失和信息泄露，用于测试智能体在进化过程中的韧性。2. **基于大语言模型的自动化表型分析：**我们开发了一套包含四项测试（圣徒测试、挑衅测试、噪声容忍度测试、贪婪测试）的行为分析工具，用于量化智能体心理；随后利用 GPT-5.1 将这些指标综合为叙事性角色档案。3. **冠军名人堂：**我们实现了一个持久化存储系统，使不同场次中进化出的冠军能够相互参加锦标赛。

1.3 相关工作与我们的创新

近期研究探讨了将大语言模型用作迭代囚徒困境中的博弈智能体，并有研究表明，GPT-4 等模型表现出的合作行为优于人类基准 [4, 1]。Park 等人 [9] 提出了“生成式智能体”，通过大语言模型驱动的反思与规划来模拟可信的人类行为。然而，这些方法将大语言模型用作 决策智能体，而不是用来解读进化智能体。

T能够 1总结我们的框架与相关方法之间的区别 .

表 1：我们的框架与相关方法的比较

方面	传统遗传算法迭代囚徒困境	大语言模型作为玩家	我们的框架
策略来源	进化基因型	大语言模型提示	进化基因型
环境	确定性	确定性	随机（上帝模式）
可解释性	无（仅二元）	隐含于大语言模型中	大语言模型生成的画像
行为测试	事后分析	N/A	集成式分析器
字符输出	None	N/A	名称、格言、阵营

2 方法论

2.1 基因组表示

每个智能体的策略都编码为一个 18 位基因型，用于实现一个 2 轮查找表（LUT）。这种方法源自早期的演化迭代囚徒困境研究 [5]。基因组结构定义如下：

- **基因 0**：第一次行动（没有可用的历史记录）
- **基因 1**：第二次行动（1 轮历史记录）
- **基因 2-17**：两轮历史的全部 16 种组合的响应表

For 对于给定的历史记录，通过计算基因型中的索引来确定响应 pe:

$$Index_t = 2 + \sum_{i=1}^{k} S_{t-i} \cdot 4^{i-1} \tag{1}$$

其中， S_{t-i} 编码第 $t-i$ 轮的结果：CC $= 0$、CD $= 1$、DC $= 2$、DD $= 3$。对于 $k = 2$（记忆深度为 2 轮）：

$$Index = 2 + (Outcome_{t-2} \times 4) + Outcome_{t-1} \tag{2}$$

响应为 $A_t = G[Index] \in \{0, 1\}$，其中 $0 =$ 表示背叛， $1 =$ 表示合作。

2.2 进化算子

遗传算法采用四种标准算子，使种群在若干代中不断进化 [6]。

选择（轮盘赌）： 根据适应度按比例选择亲本：

$$P(agent_i) = \frac{fitness_i}{\sum_{j=1}^{N} fitness_j} \tag{3}$$

交叉（均匀）： 给定两个亲本基因型 G_A 和 G_B，通过以下方式生成子代基因型：

$$G_{child}[k] = \begin{cases} G_A[k] & \text{with probability } 0.5 \\ G_B[k] & \text{with probability } 0.5 \end{cases} \quad \forall k \in [0, 17] \tag{4}$$

变异（位翻转）： 每个基因可能以概率 μ发生翻转：

$$G[k] \leftarrow 1 - G[k] \quad \text{with probability } \mu \tag{5}$$

我们使用 $\mu = 0.02$ （每个基因2%）。

Eli精英保留： 种群中排名靠前的 e 比例会被直接复制到下一代 :

$$Population_{t+1}[0 : eN] = \text{top } eN \text{ agents from } Population_t \tag{6}$$

我们使用 $e = 0.1$ （10%的精英保留）。

2.3 随机环境压力因素

我们定义了五个相互独立的环境压力因素（在实现中称为“上帝模式”），每个因素以概率 p 在每轮次或每个智能体上施加影响。该方法扩展了此前关于噪声迭代囚徒困境环境的研究 [12]。

应激因子 1：颤抖之手。 以 ϵ 的概率，智能体预定的行动会被反转 :

$$A_{actual} = \begin{cases} 1 - A_{intended} & \text{with probability } \epsilon \\ A_{intended} & \text{with probability } 1 - \epsilon \end{cases} \tag{7}$$

应激因子 2：经济危机。 以 p_{crisis} 的概率，所有收益都会按 $\alpha < 1$ 缩放：

$$Payoff_{crisis} = \alpha \cdot Payoff_{standard} \tag{8}$$

应激因子 3：高诱惑。 以 p_{greed} 的概率，诱惑收益会乘以 $\beta > 1$：

$$T_{modified} = \beta \cdot T_{standard} \tag{9}$$

应激因子 4：记忆丧失。 以 $p_{amnesia}$ 的概率，智能体感知到的历史记录会被清除。

应激因子 5：信息泄露。 以 p_{spy} 的概率，智能体会观察到对手的预定行动。

2.4 行为分析器

进化后，冠军智能体将接受四项受控实验：

测试 A：圣徒测试 – 对手：始终合作（50 轮）。指标：背叛次数。

测试 B：挑衅测试 – 对手：挑衅者（先合作，在第 10 轮背叛，然后继续合作）。指标：恢复合作所需轮次。

测试 C：噪声容忍度测试 – 对手：以牙还牙策略，在第 10 轮被迫意外背叛。指标：能否在 2 轮内恢复。

测试 D：贪婪测试 – 对手：以牙还牙策略，在第 20–30 轮设置高诱惑（$T = 10$）。指标：背叛率的变化。

2.5 大语言模型角色生成流程

分析器结果被整合为一条供 GPT-5.1 使用的结构化提示词：

流程：原始统计数据 -> 提示工程 -> 大语言模型（GPT -5.1） ->JSON档案

大语言模型返回一个包含以下内容的 JSON对象：姓名、格言、描述、RPG_阵营。

3 实验设置

3.1 参数

表 2：实验参数

参数	值	描述
种群规模	50	每代智能体数量
代数	100	进化周期
每场比赛轮数	150	每次对手配对的迭代囚徒困境轮数
变异率	2%	每个基因的翻转概率
精英比例	10%	顶尖智能体原样复制
选择方法	轮盘赌	适应度比例
交叉	均匀	50% 逐基因继承

3.2 上帝模式概率

表 3：随机应激因子概率

应激因子	概率	效果
颤抖之手	每个代理5%	行动翻转
经济危机	每轮2%	收益 $\times$ 0.5
高诱惑	每轮10%	T: $5 \to 10$
记忆丧失	每个代理5%	历史记录清除
信息泄露	每个代理5%	查看对手的行动

3.3 收益矩阵

遵循阿克塞尔罗德的标准迭代囚徒困境收益 [2]：其中 T=5、R=3、P=1、S=0 满足 $\mathrm{T} > \mathrm{R} > \mathrm{P} > \mathrm{S}$，且 $2\mathrm{R} > \mathrm{T} + \mathrm{S}$。

表4：迭代囚徒困境收益矩阵

	C	D
C	3, 3	0, 5
D	5, 0	1, 1

4 结果与分析

4.1 系统架构

如图 1所示，该框架由三个层级组成：基于 Streamlit 的用户界面层、集成随机压力因素的仿真引擎，以及用于行为画像的分析引擎。

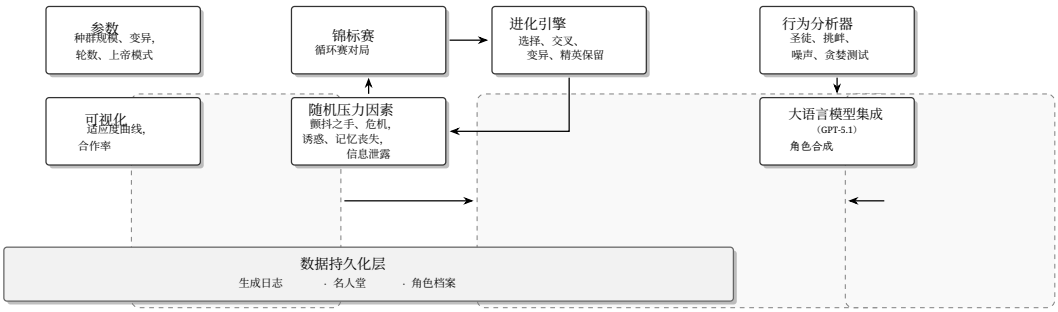


图1：演化博弈迭代囚徒困境框架概览，展示三层架构：用于可视化的 Streamlit 用户界面、集成随机应激因子的仿真引擎，以及用于行为画像和大语言模型角色生成的分析引擎。

4.2 适应度动态

图2展示了两种实验条件下 100 代的适应度轨迹。

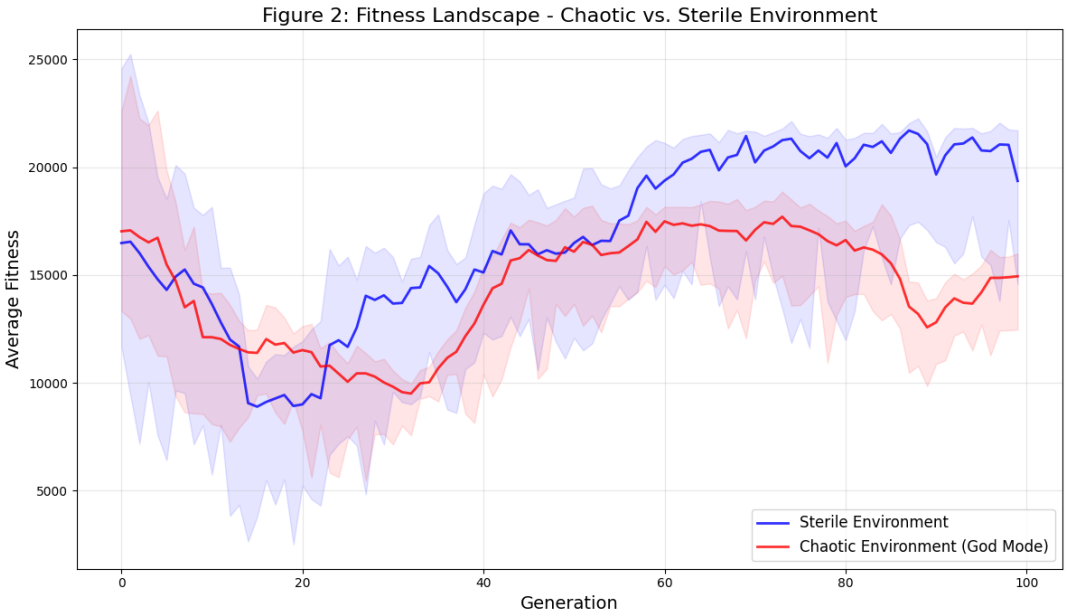


图2：100代内适应度轨迹的比较。无菌环境（蓝色）通过快速优化且不受惩罚，获得了更高的绝对适应度；而混乱环境（红色）在整个进化过程中保持较高的方差。随机压力因素施加了直接的适应度惩罚，导致平均适应度较低，但有利于筛选出能够抵御扰动的策略。

4.3 冠军案例研究：偏执的和平主义者

为展示完整的分析流程，我们对一位在随机压力条件下的后期演化中脱颖而出、成为主导者的冠军智能体进行了详细的案例研究。

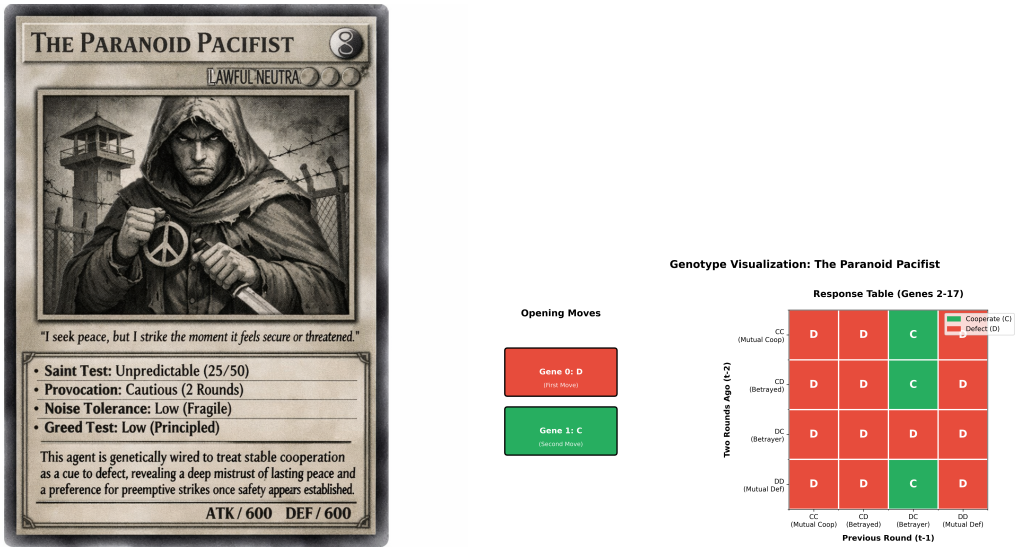


图3：“偏执的和平主义者”冠军智能体的完整档案。(a) 集换式卡牌风格的角色卡，使用 GPT-5.2 根据大语言模型合成的角色属性（姓名、格言、立场、描述）生成。(b) 基因型热图可视化，展示18基因响应模式，其中绿色表示合作，红色表示背叛。

表5：行为概况摘要：偏执的和平主义者

指标	结果	分类
基因型	先背叛，再合作；大多时候选择背叛	防御型背叛者
圣徒测试	25/50 次背叛行为	不可预测
挑衅测试	在2轮内恢复	谨慎宽恕
噪声容忍度	无法恢复	低（脆弱）
贪婪测试	-12.1% 变化率	低（有原则）
合作率	22.2%（4/18 个基因）	背叛主导

大语言模型生成的角色档案：

- **姓名**：偏执的和平主义者
- **格言**：“我追求和平，但一旦觉得安全或受到威胁，我便会立即出手。”
- **阵营**：守序中立
- **描述**：这个智能体的基因决定了它会将稳定合作视为背叛的信号，这暴露出它对持久和平的深深不信任，以及一旦安全似乎得到保障便倾向于先发制人地打击。

4.4 环境压力因素的影响

分析　　在有无随机压力因素的情况下，策略的演化呈现出显著差异　　es:

1. **纯合作者的灭绝**：与“永远合作”等价的策略在高诱惑激活时会迅速遭到利用，无法存活至第20代之后。

2. **宽恕的涌现**：颤抖之手应激因子会筛选出能够从意外背叛中恢复的智能体。

3. **对记忆干扰的适应力**：记忆丧失应激因子有利于具有合作开局动作的智能体。

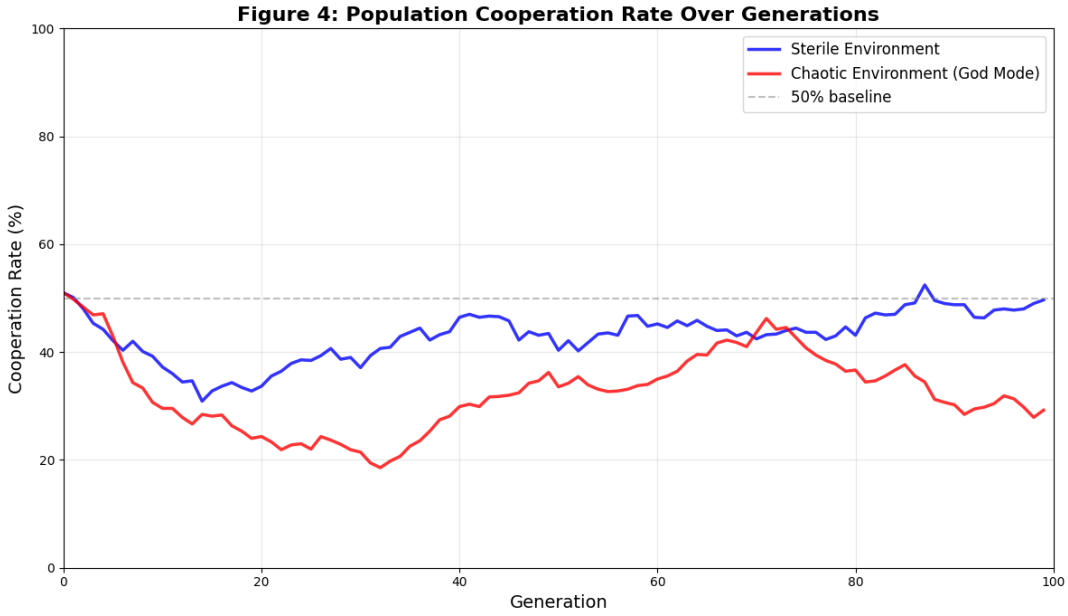


图4：100代期间群体层面的合作率。两种环境的初始合作率均约为50%。无菌环境（蓝色）将合作率维持在接近50%基线的相对稳定水平。混乱环境（红色）则先是显著下降（第0–35代），随后部分恢复（第35–70代），之后再次下降——这表明持续的环境压力会继续筛选出背叛主导策略。

4.5 大语言模型角色档案生成

GPT-5.1模型成功为所有受测冠军生成了连贯的角色档案。定性分析显示：

- **名称恰当性**：大语言模型生成的名称准确反映了智能体的主导策略。
- **阵营一致性**：角色扮演游戏风格的阵营与行为指标表现出定性一致性。
- **新颖洞见**：LLM描述识别出了原始指标未能立即显现的细微策略模式。

5 讨论

5.1 弥合可解释性鸿沟

我们的结果表明，大语言模型可以充当进化智能体的有效“表型翻译器”。结构化提示工程方法为智能体解释提供了一套可复现的流程。

这种方法解决了遗传算法的一个根本局限：优化成功与人类理解之间的脱节 [3]。

5.2 上帝模式作为进化压力

随机压力因素具有双重作用：

1. **生态效度**：现实世界的策略环境涉及不确定性、沟通失误和不断变化的激励因素。2. **进化选择**：通过在进化过程中（而非进化之后）施加压力，我们能够筛选出本质上具有稳健性的策略。

“宽恕型”策略在颤抖之手压力下的出现，映照了生物学研究的发现：宽恕机制会在嘈杂环境中演化 [8]。

5.3 局限性与未来工作

1. **固定基因组长度**：18 个基因的结构限制了策略的复杂性。
2. **单一大语言模型**：依赖 GPT-5.1 可能会引入偏差。
3. **双人动态**：扩展到 N 人公共物品博弈将提高生态效度。

6 结论

本文提出了一个用于在迭代囚徒困境中进化出具备可解释性的智能体的综合框架。通过将随机环境压力因素与大语言模型驱动的行为分析相结合，我们同时解决了传统进化方法在生态贫瘠性和可解释性方面的局限。

我们的实验表明：

1. 环境混沌会筛选出坚韧且宽容的策略
2. 大语言模型能够将进化基因型准确地转化为叙事性角色档案
3. 所得到的原型为理解复杂的策略行为提供了直观方式

参考文献

[1] Akata, E., Schulz, L., Coda-Forno, J., Oh, S. J., Bethge, M., & Schulz, E.（2023年）。使用大型语言模型进行重复博弈。*arXiv* 预印本 *arXiv:2305.16867*。

[2] Axelrod, R.（1984）。合作的演化。Basic Books。

[3] Barredo Arrieta, A. 等（2020）。可解释人工智能（XAI）：面向负责任人工智能的概念、分类体系、机遇与挑战。*Information Fusion*，58，82-115。

[4] Brookins, P. 与 DeBacker, J. M.（2023）。与 GPT 玩博弈：我们能从规范战略博弈中了解大语言模型的哪些信息？*arXiv* 预印本 *arXiv:2305.07970*。[5] Fogel, D. B.（1993）。迭代囚徒困境中的行为进化。《进化计算》，1(1)，77-97。[6] Hofbauer, J. 与 Sigmund, K.（1998）。《进化博弈与种群动力学》。剑桥大学出版社。[7] Lindgren, K.（1991年）。简单动力学中的演化现象。载于《人工生命 *II*》（第295-312页）。Addison-Wesley。[8] Nowak, M. A. 与 Sigmund, K.（1992年）。异质种群中的以牙还牙。《自然》，355(6357)，250-253。[9] Park, J. S. 等（2023）。生成式智能体：人类行为的交互式模拟体。载于 *UIST '23*。ACM。[10] Selten, R.（1975年）。重新审视均衡点的完全性概念。《国际博弈论杂志》，4(1)，25-55。[11] Virgolin, M. 等（2021年）。改进用于符号回归的基于模型的遗传编程。《进化计算》，29(2)，211-237。[12] Wu, J. 与 阿克塞尔罗德, R.（1995年）。如何应对迭代囚徒困境中的噪声。《冲突解决杂志》，39(1)，183-189。

基因型解码示例

考虑基因型 `[1, 1, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1]`:

Gene	索引范围	含义
基因 $0 = 1$	-	第一步：合作
基因 $1 = 1$	-	第二步：合作
基因 2-5	在 CC 之后	[1,0,0,1]
基因 6-9	在 CD 之后	[1,0,0,1]
基因 10-13	DC 之后	[0,1,1,0]
基因 14-17	DD 之后	[1,0,0,1]

这编码了一个宽容的以牙还牙变体。

B 随机压力事件日志样本

```
[事件 ]R47 |A3 对阵 A12: [危机 ]经济危机——收益减半[事件 ]R48 |A3 对阵 A12: [
```

代码可用性： 源代码可在https://github.com/Oguzhanyldrmm/Adaptive-Prisoner