arXiv:2611.30200v1 [cs.AI] 23 Nov 2026

规避陷阱地采用人工智能技术：一种多智能体强化学习框架

Jitong Yang* Avinash Reddy Mudireddy† Ziqing Lu ‡ HuanCai § Weiyu Xu ¶‖

2026年11月24日

摘要

人工智能技术和人工智能智能体在人类社会中发挥着日益重要的作用，包括教育、经济学和研究领域。采用人工智能技术可以带来即时、有时甚至巨大的奖励，例如提升生产力，但也可能造成负面影响，包括但不限于失去工作和培训机会，以及认知能力下降。因此，研究如何以最优方式采用人工智能技术、避免落入各种陷阱至关重要。本文提出一个基于多智能体马尔可夫决策过程（MDP）的强化学习（RL 或多智能体强化学习）通用建模框架，用于确定并分析人工智能技术的最优采用方式。该框架能够对人工智能应用的长期动态影响进行建模，并据此找到最优人工智能采用策略。在我们的建模框架中，人工智能应用的负面影响通过MDP中的不良状态转移或负奖励来建模。所提出的多智能体强化学习MDP同时纳入人类智能体和人工智能智能体。我们模型的核心思想是，采用人工智能技术不仅可能增加人工智能智能体的即时奖励，还可能改变相关人类智能体和人工智能智能体的状态。具体而言，我们运用该框架分析人工智能卸载如何导致技能退化、采用人工智能的企业如何优化人力资源结构，以及人类智能体与人工智能智能体之间交易和就业的经济均衡。由于基于多智能体MDP的强化学习技术源自控制、机器学习和人工智能领域，这项研究本质上是运用人工智能技术，确定如何以最优方式采用人工智能或与人工智能共存，同时避免陷阱。

*爱荷华大学电气与计算机工程项目。
†爱荷华大学电气与计算机工程项目。
‡维克森林大学计算机科学系，北卡罗来纳州温斯顿-塞勒姆。§康奈尔学院经济与商学系，爱荷华州弗农山，邮编 52314。
¶爱荷华大学电气与计算机工程系，爱荷华城，爱荷华州 52242。电子邮箱：weiyu-xu@uiowa.edu。
‖作者按字母顺序排列。

1 引言

人工智能技术及相关智能体在人类社会中发挥着日益重要的作用，包括教育、经济学和研究领域（Goldfarb and Tucker，2019年；Abrardi, Cambini, and Rondi，2022年；Agrawal 等，2023；Kapoor 等，2024；Li 等，2024）（Hammond 等，2025）Cai、Xu 和 Xu（2023）Bauer（2026年）Aouad, Lykouris, and Zhong（2026年）Dasaklis 等（2026年）Akin（2026年）Kimura（2026年）。采用人工智能技术可以带来即时、有时甚至巨大的奖励，例如提升生产力，但也可能造成负面影响，包括但不限于失去工作、失去培训机会以及认知能力下降。因此，研究如何以最优方式采用人工智能技术、避免落入各种陷阱至关重要。

采用人工智能技术除了能带来即时奖励外，也可能产生负面影响，但这些负面影响未必会立即显现，而可能要到未来才会出现。因此，要以最优方式采用人工智能技术，就不能只关注短期回报，还要重视采用人工智能技术所带来的长期回报。然而，由于采用人工智能技术是相对较新的现象，人们也没有采用人工智能的最佳方式方面的经验，因此，人们很容易为了获得最大的即时奖励而短视地采用人工智能技术。这种短视的人工智能应用，因其既会取代人类劳动力和思维，也会增强人类劳动力和思维，可能导致极不理想的未来后果。由此，我们面临着一个问题：如何才能以最优方式采用人工智能技术，从而获得可持续的长期回报？

本文构建了一个通用的、基于多智能体马尔可夫决策过程（MDP）的强化学习（RL 或多智能体强化学习）框架，用于确定和分析人工智能技术的最优应用方式。该框架能够对人工智能应用的动态长期影响进行建模，并据此找出最优人工智能采用策略。在我们的建模框架中，人工智能应用的负面影响通过MDP中的状态转移或负奖励来建模。所提出的多智能体强化学习框架纳入了人类智能体和人工智能智能体，并可利用MDP模型的策略来刻画这些智能体之间的经济均衡。

我们研究人工智能应用的方法，其核心思想是：引入人工智能技术和人工智能智能体，不仅可能提高人工智能智能体的即时奖励，还可能改变环境以及其中人类和人工智能智能体的状态。在经济中，采用人工智能可能影响人类和人工智能智能体的就业机会，并改变智能体之间的财富分配，从而导致智能体处于不同的财富状态或就业状态。例如，如果某个社区将所有经济任务都交给成本更低、却不会促进该社区财富增长的外部人工智能智能体，该社区可能因此失去就业机会和增长机会。对于企业而言，使用人工智能技术取代初级员工可能导致劳动力准备不足：当企业面临资深员工流失或其他未来新颖挑战时，

企业可能会陷入功能失调。例如，企业可能会受到诱惑，只让资深员工与人工智能合作，以较低的总成本完成工作任务。然而，当资深员工离职或退休时，这会导致企业陷入功能失调，因为没有初级员工在入门级岗位上接受过训练。对于幼儿园至高中和大学生而言，使用人工智能完成家庭作业可以带来即时奖励，例如课程成绩更高，但也可能导致学生对所学科目的理解不足。对于数学家而言，使用人工智能智能体直接生成人工智能证明，可以获得快速解决数学猜想或问题的奖励；但与此同时，这会使初级数学家失去培训机会，也可能使数学家失去在独立解决数学猜想或问题时，通过健康的迂回探索取得新发现和新理解的机会。对于员工而言，将任务卸载给人工智能会影响他们的技能水平：虽然他们可以立即获得更高的生产力，但任务卸载也会损害其技能，使他们在出现无法由人工智能充分处理的新颖任务时表现不佳。

此外，在由多个人工智能智能体和人类智能体组成的系统中，所有智能体都可以决定如何采用人工智能技术，而它们的联合行动会影响环境及智能体自身的状态。因此，如何以最优方式采用人工智能技术这一问题，属于多智能体马尔可夫决策过程问题，而非单智能体马尔可夫决策过程问题。

本文聚焦于三种应用，并推导出相应的人工智能应用最优策略：工作卸载与认知技能训练；人工智能的使用与劳动力发展；人类智能体与人工智能智能体之间的交易。从短视的角度看，为了最大化即时奖励，员工将任务卸载给人工智能似乎总是最优的，因为人工智能能更高效地完成任务；对企业而言，只保留与人工智能协作时生产力较高的资深员工似乎总是最优的，因为初级员工可以被人工智能取代；对人类智能体而言，只从人工智能智能体处购买产品/服务似乎总是最优的，因为人工智能能以更低的成本提供更好的产品/服务。然而，盲目采用人工智能会改变环境及相关人类的状态（技能水平、劳动力状况和财富分配），因此未必总能使整体或长期奖励最大化。我们的多智能体马尔可夫决策过程形式化为确定人工智能技术的最优采用方式提供了一个有原则的框架，既能最大化总体奖励，又能避免陷阱。

本文结构如下：第3节介绍采用人工智能技术的多智能体强化学习框架的数学模型。第4节介绍关于人工智能卸载和技能使用的马尔可夫决策过程。第5节介绍企业采用人工智能并储备劳动力的马尔可夫决策过程。第6节介绍同时包含人类智能体和人工智能智能体的系统所对应的多智能体经济模型。最后，我们在第7节总结全文。

2 相关工作

Bondi 和 Johnson（2026）研究了人工智能的使用如何影响员工技能发展以及生产力提升的衡量。他们构建了一个动态模型，考察将任务委派给人工智能如何随时间影响员工技能。他们的研究结果表明，尽管人工智能能够提升短期生产力，但过度依赖人工智能可能导致技能退化。他们还发现，标准的生产力衡量方法可能会高估人工智能的长期收益，因为这些方法没有充分考虑员工技能的变化。Caosun 和 Aral（2026）研究了人工智能带来的短期生产力提升与员工长期技能发展之间的权衡。他们构建了一个动态模型，考察人工智能的使用如何随时间影响员工的生产力和技能。他们的研究结果表明，尽管人工智能能够提升即时生产力，但过度依赖人工智能可能会削弱员工技能，并导致长期生产力下降。他们还发现，人工智能应用可能会扩大高技能工人与低技能工人之间的技能差距。

我们注意到，Bondi 和 Johnson（2026年）Caosun 和 Aral（2026年）中的动态模型主要局限于分析使用人工智能的员工的技能退化。我们的马尔可夫决策过程模型更为通用，可应用于其他情形，例如研究在许多其他方面采用人工智能所带来的影响，包括交易和就业Cai、Xu 和 Xu（2023）Cai 等（2026）Falk 和 Tsoukalas（2026）、数学家的成长以及K-12教育。在通过AI卸载研究技能退化方面，与本文不同，Bondi 和 Johnson（2026年）Caosun 和 Aral（2026年）并未考虑员工的动作直接依赖于马尔可夫决策过程（MDP）状态的情形，而仅让动作本质上只依赖于时间变量或索引。因此，Bondi 和 Johnson（2026年）Caosun 和 Aral（2026年）中的动态模型是确定性动态系统，其中状态与动作是时间的确定性函数，而不是马尔可夫随机过程。与Caosun 和 Aral（2026年）Bondi 和 Johnson（2026年）相比，我们的马尔可夫决策过程（MDP）框架允许在不同动作下进行随机且更丰富的状态转移。换言之，就本文而言，智能体的动作不仅取决于时间索引，还取决于当前马尔可夫决策过程的状态；即使对于给定的时间索引，该状态也可能取不同的可能值。事实上，在采用人工智能时，许多因素都可能是随机的。例如，工人在某一时间索引需要解决或从事的问题类型可能是随机的，取决于随机的外部应用需求。作为随机性的另一例，员工可能因退休或搬迁而随机离开或加入企业。总之，本文的马尔可夫决策过程框架比Caosun 和 Aral（2026年）Bondi 和 Johnson（2026年）中的动态系统模型更为通用。

在 Cai、Xu 和 Xu（2023）以及 Cai 等（2026）的研究中，作者没有考察不同人工智能智能体之间的短期经济相互作用，而是采用基于马尔可夫链平稳分布的模型，研究人类与 AI 智能体之间的长期经济相互作用。然而，Cai、Xu 和 Xu（2023）以及 Cai 等（2026）中的方法属于一种特殊的、不同于常规马尔可夫决策过程的模型（尽管从更广义的角度来看，它

可以被视为马尔可夫决策过程模型）：旧的联合状态（多个智能体的联合状态由这些不同智能体之间的财富分配来描述）与新的联合状态之间的关系，由另一个取决于交互图结构的马尔可夫转移概率矩阵描述。该矩阵仅描述这两个特定状态之间转换的内在机制，与常规马尔可夫决策过程模型中的转移核不同。在 Cai、Xu 和 Xu（2023）以及 Cai 等（2026）的研究中，动作是“支出模式”，不取决于财富分配状态，并且这些动作会确定性地产生下一时间索引对应的财富分配结果。需要注意的是，Cai、Xu 和 Xu（2023）以及 Cai 等（2026）中的购买行动包括就业行动，因为就业行动只是企业向已雇用和拟雇用的雇员购买劳动力。Falk 和 Tsoukalas Falk 和 Tsoukalas（2026）研究了人工智能自动化如何影响就业和经济福利。他们构建了一个竞争性经济模型，以考察企业如何做出自动化决策。研究结果表明，尽管人工智能自动化可以降低劳动力成本，但过度自动化可能导致失业并降低消费者需求。他们还发现，企业之间的竞争可能会促使企业过度应用人工智能，即使这会降低整体经济福利。然而，Falk 和 Tsoukalas（2026）中的模型是静态的，因为它只考虑一个时间索引；相比之下，我们论文中的方法是一种涉及多个时间索引的动态马尔可夫决策过程强化学习方法。

3 数学模型

在所考虑的马尔可夫决策过程环境中，我们将智能体群体记为 N，其中$|N| = n$。以下 5 元组可描述 n 智能体的有限马尔可夫决策过程环境：$(\mathrm{S}, \{\mathrm{A}_i\}_{i=1}^n, P, \{R_i\}_{i=1}^n, T)$，其中 T是有限时间步数，S 是所有智能体共有的状态空间，A_i 是第 i 个智能体的行动空间，其中$|\mathrm{A}_i| = A_i$，$\mathrm{A} = \prod_{i=1}^n \mathrm{A}_i$ 是联合行动空间。我们将 $P_{t\to t+1} : \mathrm{S} \times \mathrm{A} \times \mathrm{S} \to [0,1]$ 定义为所有智能体在时间步 t 和 $t+1$ 之间采取联合行动时的状态转移概率核，其中 $t = 0, 1, \ldots T-1$。我们用 $R_{i,t} : \mathrm{S} \times \mathrm{A} \to \mathbb{R}$ 表示第 i 个智能体在时间索引 t处的奖励函数，其中 $t = 1, 2, \ldots\ldots, T$。我们令时间索引 t的范围为 0 到 T。我们将时间步 t 称为从时间索引 $t-1$ 开始、到时间索引 t结束的时间间隔，其中 $1 \le t \le T$。

我们用 s_t 表示时间步 t所有智能体的联合环境状态，用 $a_{i,t}$表示第 i 个智能体在时间索引 t 采取的行动，并用 $a^*_{i,t}$ 表示第 i 个智能体在时间步 t采取的最优动作。所有 n 智能体在时间索引 t的联合行动为$a_t := (a_{1,t}, a_{2,t}, \ldots, a_{n,t})$。我们用 $\pi^*_{i,t}$表示第 i 个智能体在时间索引 t的最优策略，并且 $\pi^*_{i,t}(s_t) = a^*_{i,t}$。时间索引 t 处的确定性联合最优策略为 $\boldsymbol{\pi}^*_t = \{\pi^*_{i,t}\}_{i=1}^n$，记作 $\pi^*_i = \{\pi^*_{i,t}\}_{t=0}^{T-1}$。请注意，该策略取决于当前环境状态。

我们将 $\gamma \in [0,1]$ 记为折扣因子。我们用 $V_i^{\boldsymbol{\pi}},_t : \mathrm{S} \to \mathbb{R}$ 表示智能体 i 在时间索引 t、联合策略 $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_n)$ 引导下的价值函数，其中 $V_i^{\boldsymbol{\pi}}t, (s) = \mathbb{E}_{\boldsymbol{\pi}}\left[\sum_{k=t}^{T-1} \gamma^{k-t} R_{i,k}(s_k, \mathrm{a}_k) \mid s_t = s\right]$。智能体 i 在时间索引 t、状态—联合行动对 (s, a) 下的 Q函数 $Q_i^{\boldsymbol{\pi}},_t : \mathrm{S} \times \mathrm{A} \to \mathbb{R}$ 定义为 $Q_i^{\boldsymbol{\pi}},_t(s, \mathrm{a}) = \mathbb{E}_{\boldsymbol{\pi}}\left[\sum_{k=t}^{T-1} \gamma^{k-t} R_{i,k}(s_k, \mathrm{a}_k) \mid s_t = s, \mathrm{a}_t = \mathrm{a}\right]$。Q值与V值之间的关系为 $Q_i^{\boldsymbol{\pi}},_t(s, \mathrm{a}) = R_i,_t(s, \mathrm{a}) + \gamma \sum_{s' \in \mathrm{S}} P_{t \to t+1}(s' \mid s, \mathrm{a}) V_i^{\boldsymbol{\pi}},_{t+1}(s')$。

4 人工智能卸载与技能退化的马尔可夫决策过程模型

本节研究如何在将任务卸载给人工智能与维持员工技能之间实现平衡。我们首先介绍马尔可夫决策过程模型。

4.1 模型设置

我们假设员工有两种可能的技能水平：高水平和低水平。在每个时间步，员工需要完成（解决）的任务（问题）分为三种类型：可训练任务、不支持训练的重复任务，以及新颖任务。员工有两种可能的动作：要么完全将任务委托给人工智能处理，要么由人类自己尝试完成任务（包括人类将人工智能作为简单的辅助工具使用）。训练任务是指在完成过程中能够提升员工技能的任务（例如完成微积分证明题），因此，如果员工独立完成这些任务，其技能得到提升的概率会更高。重复任务是相对简单的任务（例如进行简单算术运算），与训练类任务相比，这类任务带来的技能提升概率较低。新颖任务是需要新颖且前所未见的解决方案和最新研究数据的困难任务，这些任务对人类和人工智能都具有挑战性：然而，高技能水平的人类员工有可能解决这些任务，从而为人类带来高额奖励；而人工智能由于缺乏训练数据或某些批判性思维能力和视角，无法完成这些任务。

无论工人的技能类型如何，都有三种可能需要处理的问题或任务。因此，我们将状态组织为一个二元元组：第一个元素表示工人的技能水平（高或低），第二个元素表示当前时间索引对应的任务类型（训练、重复、新颖）。因此，我们共有六种状态，编号为0至5。状态、技能水平与任务类型之间的对应关系见表1。

对于每种状态，工人可以在两种可能的动作之间进行选择：通过人力努力解决问题，或委托给人工智能。人类动作标记为0，人工智能动作标记为1。

我们为该马尔可夫决策过程定义概率转移核 $P \in \mathbb{R}^{6 \times 2 \times 6}$，它表示在两种不同动

作下，马尔可夫决策过程在不同状态之间转移的概率。 $P[i,j,:]$ 是一个 6 维概率向量，表示转移到 6 个后续状态的转移概率，其中 i ($0 \le i \le 5$) 表示当前状态， j ($0 \le j \le 1$) 表示当前动作。令 p、 q 和 r 分别表示在每个时间索引遇到训练、重复和新颖任务类型的概率；并且 $p+q+r=1$。我们假设任务类型与工人的技能水平无关，并且各时间索引上的任务类型相互独立。我们还定义奖励矩阵 $R \in \mathbb{R}^{6\times 2}$，表示工人在完成任务后获得的奖励。这里， $R(i,\ j)$ 表示与状态 i 和行动 j 相关的奖励。

例如， $P[0,0,:]=(0.7p,\ 0.7q,\ 0.7r,\ 0.3p,\ 0.3q,\ 0.3r)$ 表示低技能工人使用人力努力处理训练类问题后转移到 6 个不同状态的概率。工人以 0。7 的概率保持在低技能水平，并以 0。3 的概率转移到高技能水平。无论工人的技能水平如何，遇到三类问题的概率均由 p、 q 和 r 决定。

状态		行动	速率 ρ	转移核 P						R 矩阵
状态编号	状态详情			p	q	r	p	q	r	
状态0	低技能、训练	人类	$\rho(S_0,H)=0.7$	0.7p	0.7q	0.7r	0.3p	0.3q	0.3r	$R(S_0,\ H)=10$
状态0	低技能、训练	AI	$\rho(S_0,\ A)=1$	1p	1q	1r	0p	0q	0r	$R(S_0,A)=15$
状态 1	低，重复	人类	$\rho(S_1,H)=0.95$	0.95p	0.95q	0.95r	0.05p	0.05q	0.05r	$R(S_1,H)=9$
状态 1	低，重复	AI	$\rho(S_1,A)=1$	1p	1q	1r	0p	0q	0r	$R(S_1,A)=14$
状态 2	低、新颖	人类	$\rho(S_2,\ H)=0.8$	0.8p	0.8q	0.8r	0.2p	0.2q	0.2r	$R(S_2,H)=0$
状态 2	低，新颖	AI	$\rho(S_2,A)=1$	1p	1q	1r	0p	0q	0r	$R(S_2,A)=1$
状态 3	高，训练	人类	$\rho(S_3,\ H)=0.01$	0.01p	0.01q	0.01r	0.99p	0.99q	0.99r	$R(S_3,H)=11$
状态 3	高，训练	AI	$\rho(S_3,A)=0.05$	0.05p	0.05q	0.05r	0.95p	0.95q	0.95r	$R(S_3,\ A)=16$
状态 4	高，重复	人类	$\rho(S_4,H)=0.03$	0.03p	0.03q	0.03r	0.97p	0.97q	0.97r	$R(S_4,H)=9.5$
状态 4	高，重复	AI	$\rho(S_4,A)=0.1$	0.1p	0.1q	0.1r	0.9p	0.9q	0.9r	$R(S_4,A)=14.5$
状态 5	高，新颖	人类	$\rho(S_5,H)=0.001$	0.001p	0.001q	0.001r	0.999p	0.999q	0.999r	$R(S_5,H)=100$
状态 5	高，新颖	AI	$\rho(S_5,A)=0.04$	0.04p	0.04q	0.04r	0.96p	0.96q	0.96r	$R(S_5,\ A)=1$

表1： ρ 表示工人在完成任务后技能水平退化至低技能水平或保持在低技能水平的概率。我们将 $\rho \in [0,1]^{6\times 2}$ 定义为一个概率矩阵，其中第一个索引表示状态，第二个索引表示工人选择的行动。

我们指出， ρ 等示例值（如0.7、095、0.8等）可根据应用场景进行调整，以反映实际值。奖励值也可以调整。

4.2 数值结果分析

我们对表1所述的马尔可夫决策过程进行了价值迭代计算。结果见图1、图2、图3和图4。我们将折扣因子设为 $\gamma=0.9$。

在图1中，我们绘制了每个状态的最优V值随参数ρ变化的曲线，其中我们改变每个 ρ 参数的取值。这些曲线表明，最优V值取决于 ρ 参数，而这些参数决定了使用人工智能将如何影响工人的技能水平。

在图2中，我们绘制了每个状态的最优V值随奖励参数 R变化的曲线，其中我们改变了奖励参数的取值。从中可以看出，最优V值随着 $R(S_5,\ H)$显著增长；这是智能体在人类技能水平较高时执行新颖任务所获得的奖励。

在图3中，通过改变参数 ρ，我们观察每个状态下最优动作如何变化。可以看到，当智能体当前技能水平较低，且面对训练型任务（状态0）时，最优动作实际上是人类动作，尽管人工智能动作的即时奖励高于人类动作。我们还发现，当智能体当前技能水平较低，但面对“重复”型任务（状态 1）时，智能体会采用人工智能动作，因为与人类动作相比，人工智能动作能提供更高的奖励，同时不会显著降低工人成为高技能工作者的概率。我们还观察到，对于重复任务，最优动作通常是人工智能动作，因为人类动作提供的即时奖励较低，也不会提高工人成为高技能工作者的概率。

在图4中，我们改变奖励参数的取值，并观察最优动作如何变化。可以看到，当工人技能水平较低，且智能体面对训练型任务时，在很大范围的奖励参数取值下，智能体的最优动作都是采用人类动作，即使相应的人工智能动作能带来更多即时奖励。相反，当工人已经具备较高技能水平（状态 3）时，如果我们降低人类动作的奖励 $R(S_3,\ H)$，或提高人工智能动作的奖励 $R(S_3,\ A)$，工人就会从人类动作转向人工智能动作。这是因为通过人类动作执行任务来提升智能体技能水平所带来的收益较小，因此智能体更愿意采取人工智能动作。

我们进一步绘制了图5和图6，分别对应图1和图3的扩展，并将奖励 $R(S_5,\ H)$从100提高到1000。这意味着高技能工作者面对新颖任务时会获得更多奖励，奖励金额从100提高到1000。当完成新颖任务（例如进行一项发明）能够带来显著更多奖励时，这样的设定是合理的。从这两张图中，尤其是图6中，我们可以看到，在不同状态下，员工更倾向于采取人类动作而非人工智能动作：这是因为人类动作能够更好地保留/训练工人的技能，使得技能娴熟的智能体每当面对新颖任务时都能获得更多奖励。

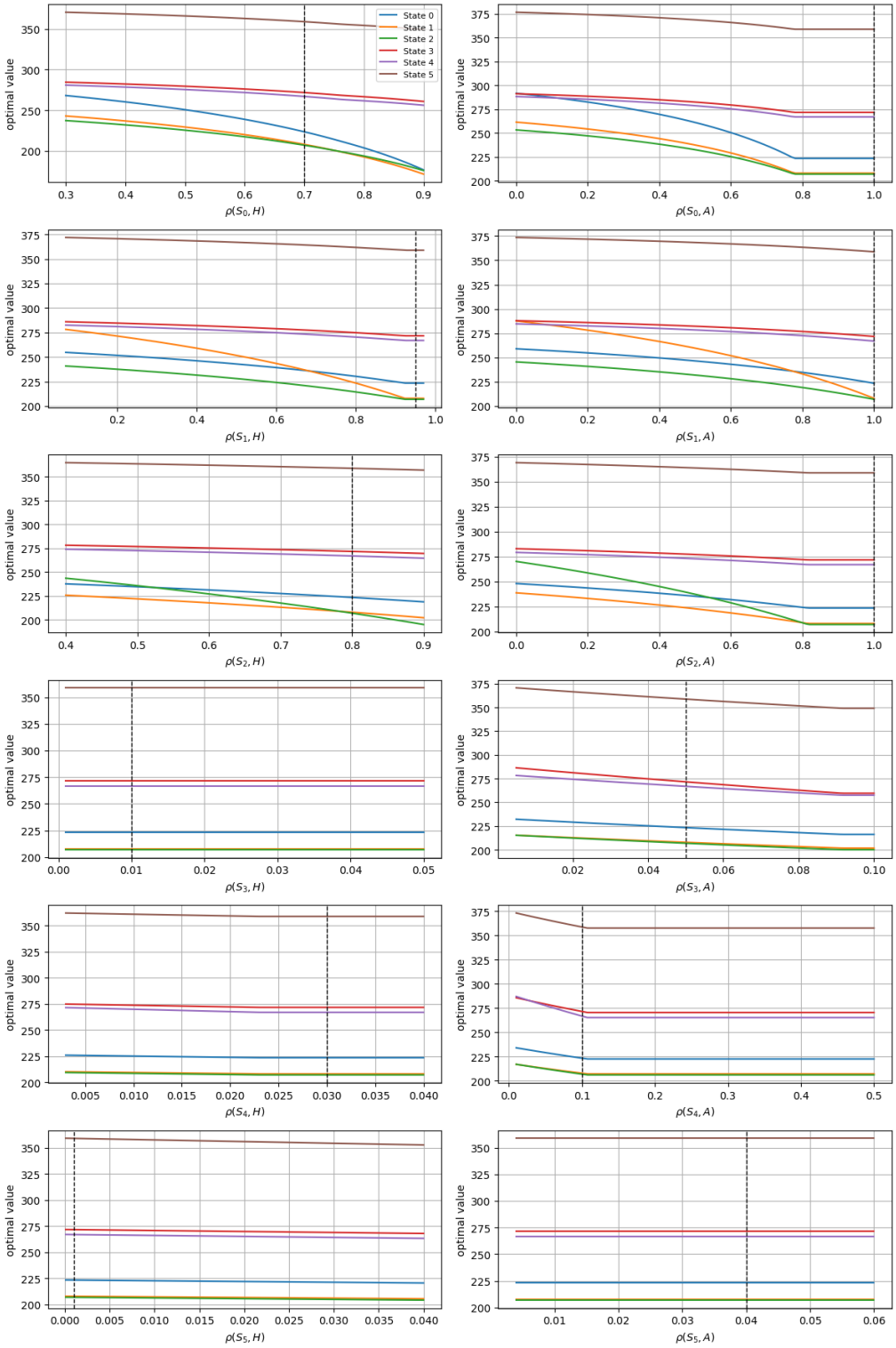

图1：最优值与 ρ的关系。该图展示了各速率概率变化时最优值如何变化。

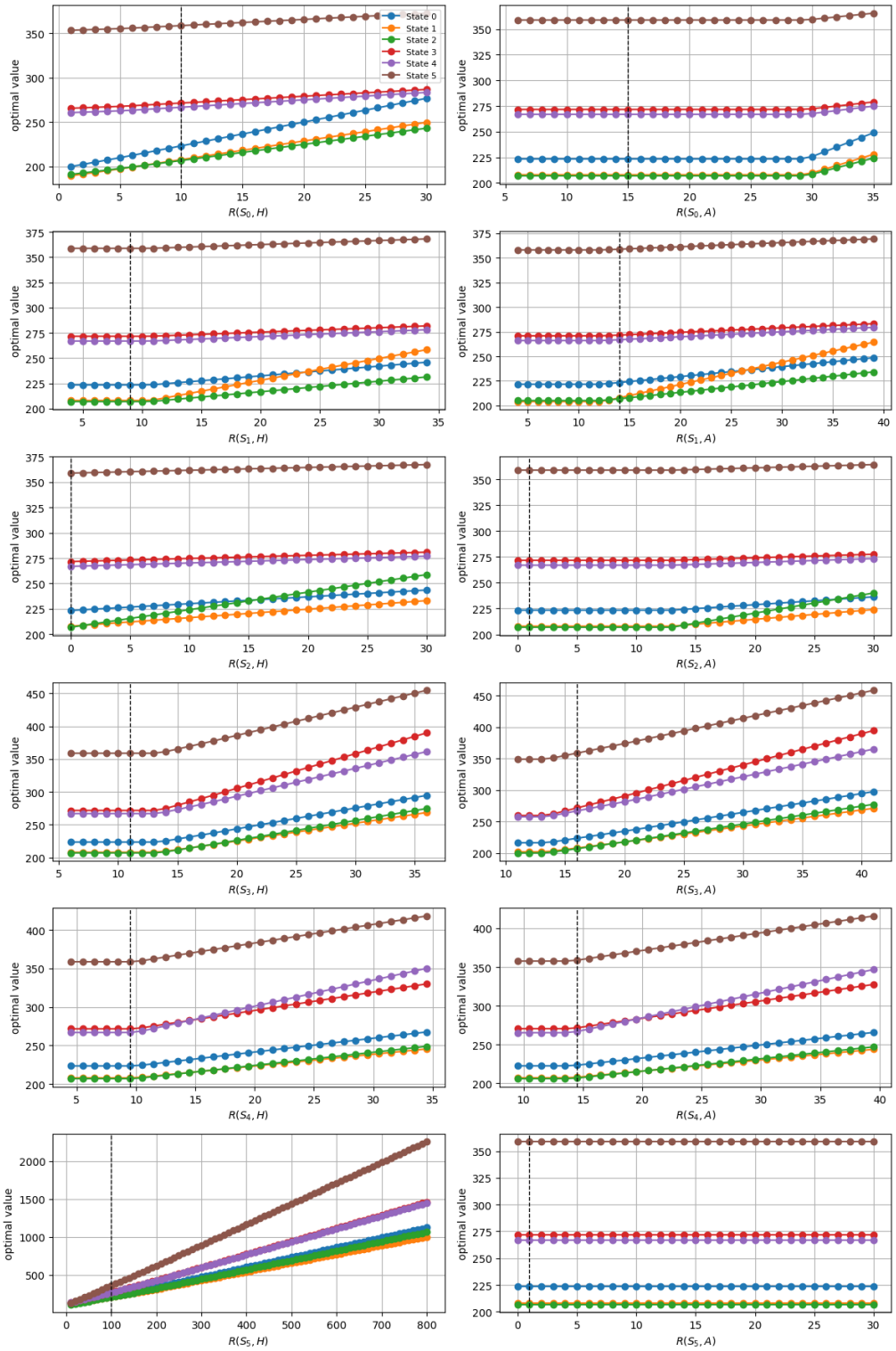

图2：最优值与 R 的关系。该图展示了每个奖励值变化时最优值的变化情况。

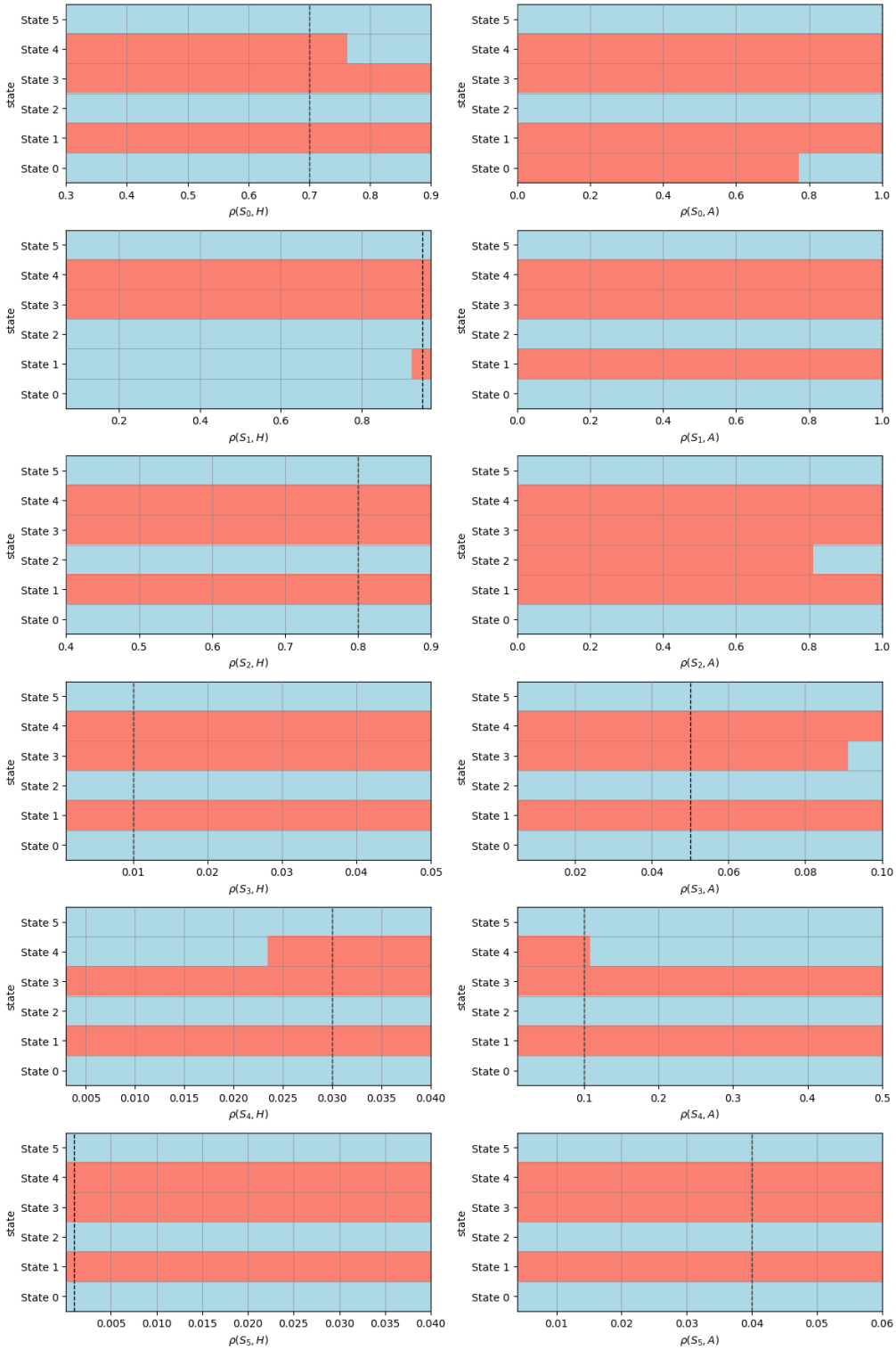

图3：最优动作与 ρ 的关系。该图展示了最优动作如何随 ρ 参数变化。对于每个状态，浅蓝色表示工人选择人类动作，而鲑红色表示工人选择人工智能动作。

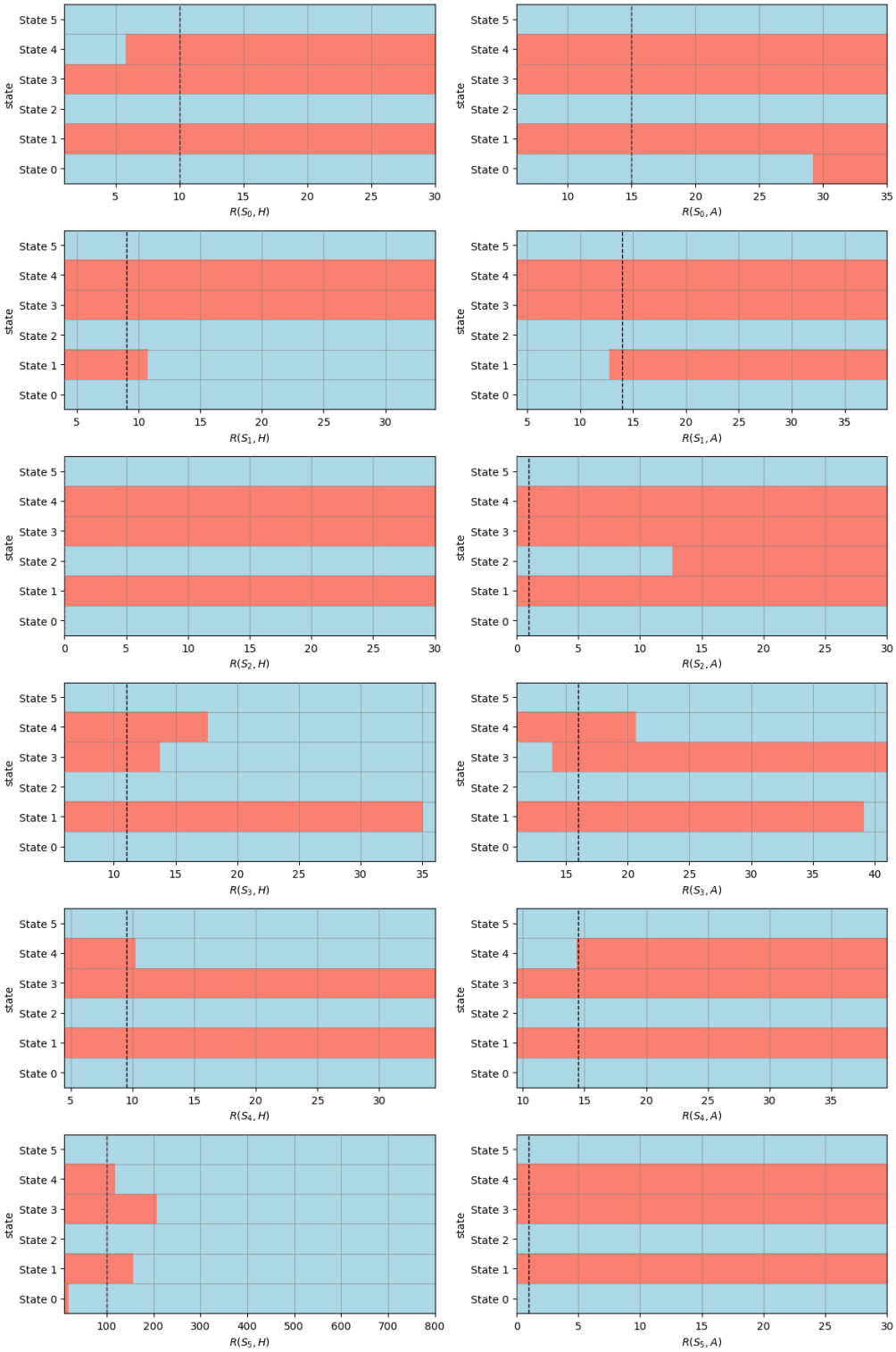

图4：最优动作与 R。此图展示了最优动作如何随各奖励值变化。浅蓝色表示工人选择人类动作，鲑红色表示工人在每种状态下选择人工智能动作。

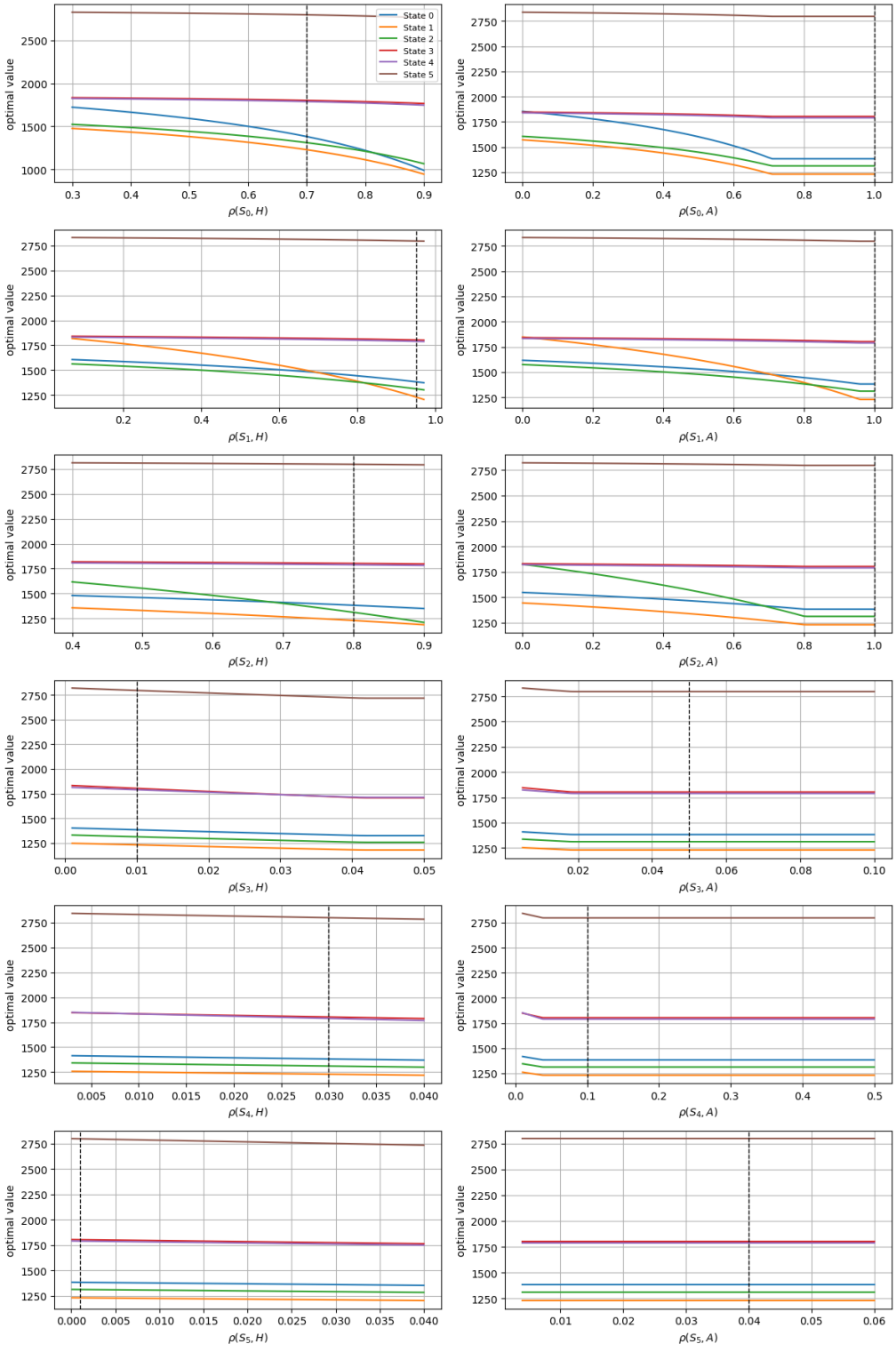

图5：奖励为1000时的最优值与 ρ 的对比。

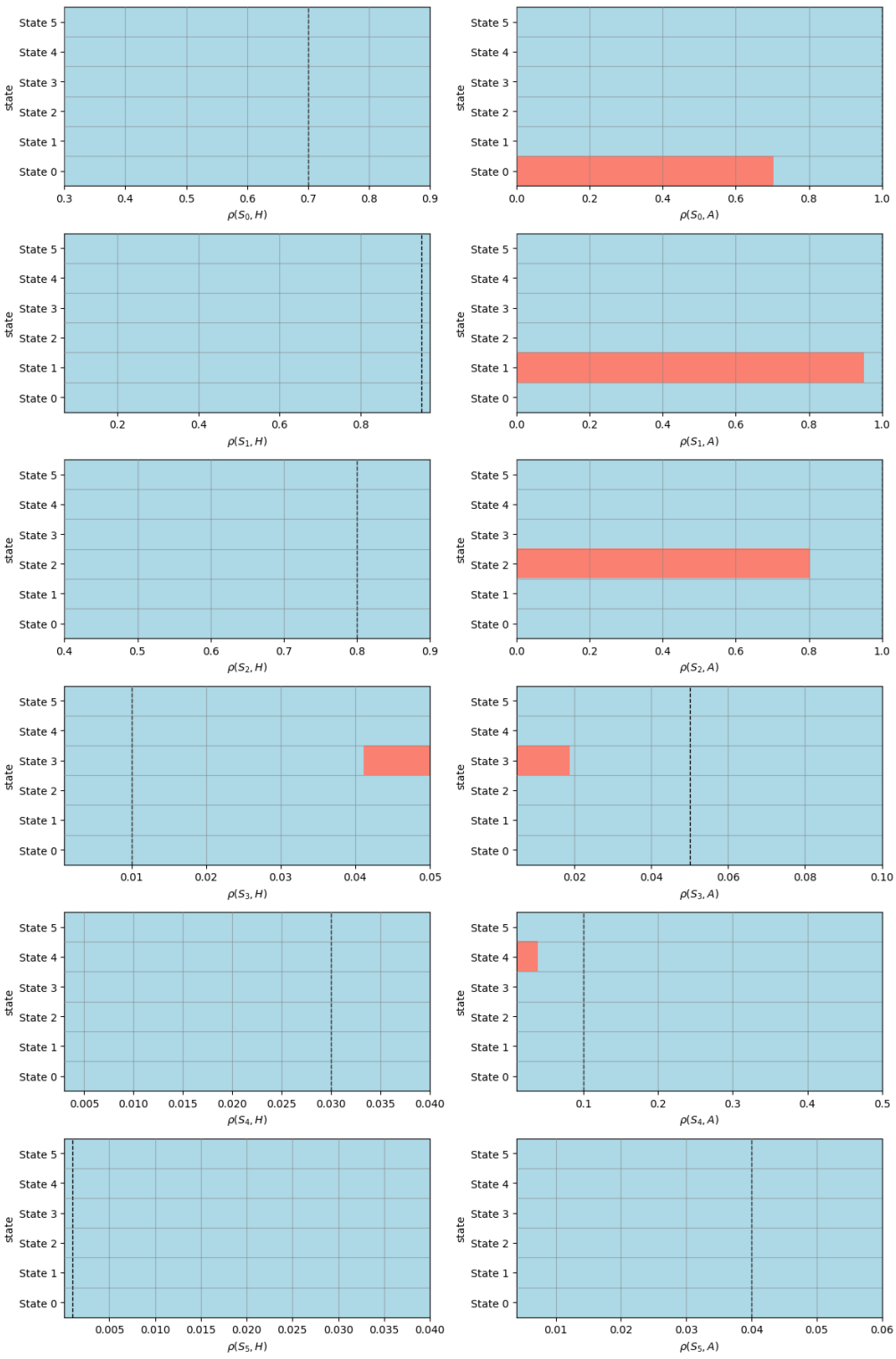

图6：奖励为1000时的最优动作与 ρ 对比。

5 企业人工智能应用与人力资源积累的马尔可夫决策过程模型

本节探讨企业应用人工智能和建设劳动力的最优方式。表面上看，只保留资深员工似乎颇具吸引力，因为他们可以借助人工智能完成任务，从而节省招聘初级员工的成本。然而，我们的马尔可夫决策过程模型表明，同时优化人工智能的使用和初级员工的招聘十分重要。

在该模型中，我们考虑一个马尔可夫决策过程，其状态由一个三元素向量表示（S，J，R）。其中，S 表示企业是否有资深员工：若有，则为 $S = 1$；否则为 $S = 0$。J 表示企业是否有初级员工：若有，则为$J = 1$；否则为 $J = 0$。R 表示初级员工（若有）是否已准备好执行资深员工能够处理的任务：若已准备就绪，则为 $R = 1$；否则为 $R = 0$。我们不允许状态（0，1，1）出现，因为在这种情况下，该状态可以立即转变为（1，0，0）。因此，总共有5种可能状态：（1，0，0）、（1，1，0）、（1，1，1）、（0，1，0）和（0，0，0）。

我们假设劳动力市场上没有已经接受培训、专门具备该企业所需领域专长的资深员工。因此，企业只能从劳动力市场招聘初级员工；此外，在我们的模型中，企业只招聘员工，不解雇员工。

我们使用另一个三维向量（SA，JA，EA）来表示企业可能采取的行动。SA 表示资深员工是否采用人工智能：如果采用，则 $SA = 1$；否则，$SA = 0$。JA 表示初级员工是否采用人工智能：如果采用，$JA = 1$；否则，$JA = 0$。EA表示公司是否招聘初级员工：如果 $EA = 1$，则表示公司进行招聘；如果 $EA = 0$，则表示公司不进行招聘。如果公司进行招聘，而现有初级员工没有晋升为资深员工，那么招聘不会增加新的初级员工。此外，只有在资深员工离职，且初级员工已准备好并能够承担资深员工的任务时，初级员工才会成为资深员工。

在该模型中，我们假设，当有空缺可供招聘初级员工时（例如，原有的初级员工已晋升并占据资深员工的职位），招聘行动以 0.91 的概率成功，以 0.09 的概率失败。招聘本身会带来 −2的奖励。如果尚未提升至资深员工水平的初级员工使用人工智能，他们会在下一个时间索引达到资深员工水平，概率为 0.3；如果尚未提升至资深员工水平的初级员工使用人工智能，他们达到资深员工水平的概率则为 0.35。如果没有资深员工，采用人工智能的初级员工会产生 8 的奖励，而不采用人工智能的初级员工会产生 6 的奖励。但保留初级员工也会因劳动力成本而使奖励减少 3。资深员工在每个时间步内离开劳动力的概率为 0.93。当资深员工采用人工智能时，企业

无论初级员工采取何种行动，都会产生 30 的奖励。当资深员工不采用人工智能时，他们会产生 10 的奖励。但我们不单独考虑留用资深员工的劳动力成本，因为我们已从奖励中扣除了该成本（即扣除成本后的奖励为 30 或 10）。我们还假设，当资深员工使用人工智能时，初级员工不会对产生任何奖励作出贡献，因为在资深员工的监督下，人工智能可以替代初级员工的劳动贡献。我们指出，这些数值（0.91、0.93、30、10、3、2）可以推广为可取其他值的一般变量。

根据这些规则，我们得到相应马尔可夫决策过程的以下转移概率和奖励结构。

5.1 状态 (0,0,0)

假设当前状态为 (0，0，0)，即该公司中没有高级员工或初级员工。

如果行动为 (SA，JA，1)，其中 SA和 JA可取任意值，则状态将以 0。91 的概率转移到 (0，1，0)，并以 0。09 的概率转移到 (0，0，0)。这些转移的奖励为 -2。

如果行动为 (SA，JA，0)，其中 SA和 JA可取任意值，则马尔可夫决策过程的状态将以概率 1 转移到 (0，0，0)，且此转移产生的奖励为 0。

5.2 状态 (0,1,0)

假设当前状态为 (0,1,0)，也就是说，这家公司只有初级员工。

5.2.1 行动 (SA,1,1)

如果行动为 (SA,1,1)，其中 SA 可以取任意值，则初级员工会采用人工智能，企业会进行招聘。资深员工是否采用人工智能，取决于 SA 的取值。

随后，状态将以 0.7 的概率转移至 (0,1,0)，奖励为 $8-3-2=3$；其中 -3 是初级员工劳动力成本所导致的扣减，-2 是招聘所获得的奖励。

状态将以 0.3×0.91 的概率转移至 (1,1,0)，奖励为 $8-3-2=3$；其中 0.3 是初级员工达到资深员工技能水平的概率，0.91 是成功招聘的概率。

状态将以概率 0.3×0.09 转移到 (1,0,0)，并获得奖励

$8-3-2=3$，其中 0.3 是初级员工达到资深员工技能水平的概率，而 $0.09=1-0.91$ 是招聘失败的概率。

5.2.2 行动 (SA,1,0)

如果行动为 $(SA，1，0)$，其中 SA 可以取任意值，初级员工采用人工智能，企业不进行招聘。高级员工是否采用人工智能取决于 SA 的取值。

随后，状态将以 0，1， 0.7 的概率转变为 (0，1， 0)，并获得奖励 $8-3=5$。

状态将以 0，0， 0.3 的概率转变为 (1，0， 0)，并获得奖励 $8-3=5$。

5.2.3 行动 (SA,0,1)

如果行动为 $(SA，0，1)$，其中 SA 可以取任意值，初级员工不采用人工智能，企业进行招聘。高级员工是否采用人工智能取决于 SA 的取值。

随后，状态将以 0，1，0.65 的概率转移到 $(0,1,0)$，并获得奖励$6-3-2=1$；其中，$6-3=3$ 是初级员工不采用人工智能时创造的奖励，-2 是招聘所导致的扣减，而 0.65 是仍为初级员工的初级员工不采用人工智能的概率。

状态将以 $0.91*0.35$ 的概率转移到 $(1,1,0)$，并获得奖励$6-3-2=1$。

状态将以 $0.09*0.35$ 的概率转移到 $(1,0,0)$，并获得奖励$6-3-2=1$。

5.2.4 行动 (SA,0,0)

如果行动为 $(SA,0,0)$，其中 SA 可以取任意值：也就是说，初级员工不采用人工智能，且企业不进行招聘。资深员工是否采用人工智能，取决于 SA 的取值。

然后，状态将以 0，65 的概率转移到 (0，1，0)，并获得奖励$6-3=3$。

状态将以 0，35 的概率转移到 (1，0， 0)，并获得奖励 $6-3=3$。

5.3 状态 (1,1,0)

假设当前状态为 (1，1，0)，也就是说，企业中既有高级员工，也有初级员工，但初级员工尚未准备好执行只能由高级员工完成的任务

只能由高级员工执行。

5.3.1 行动 (1,1,1)

如果行动为 (1，1，1)，资深员工和初级员工都会采用人工智能，且企业会进行招聘。

随后，状态将以概率 $0.07 * (0.3) * 0.09$ (其中 0.07 是资深员工离开企业的概率) 转移到 (1，0，0)，并获得奖励 $(30-3-2)=25$。

状态将以概率 $0.93 * 0.7 + 0.07 * 0.3 * 0.91$(资深员工留任且初级员工没有进步；或资深员工离职、初级员工进步且招聘成功) 转移到 (1， 1，0)，并获得奖励 (30-3-2)$=25$。

状态将以概率 $0.93 * 0.3$(资深员工没有离职，且初级员工提升至资深级别) 转移到 (1，1， 1)，并获得奖励 $30-3-2=25$。

状态将以概率 $0.07 * 0.7$ 转移到 $(0,1,0)$（即资深员工离职，初级员工没有提升），奖励为 30-3-$2=25$。

状态将以概率 0 转移到 $(0,0,0)$，奖励为 30-3-$2=25$（其中 3 是雇用初级员工的成本，此处的奖励 30 来自资深员工使用人工智能工具工作；2 是招聘成本）。

5.3.2 行动 (1,1,0)

如果行动为 (1,1,0)，资深员工和初级员工都会采用人工智能，且企业不进行招聘。

随后，状态将以概率 (0.07)*(0.3) 转移到 (1,0,0)（其中 0.07 是资深员工离开企业的概率），奖励为 $(30-3)=27$。

状态将以概率 $0.93 * 0.7$（资深员工没有离职，且初级员工没有晋升到资深级别）转移至 $(1,1,0)$，奖励为 $30-3=27$;

状态将以概率 $0.93 * 0.3$（资深员工没有离职，且初级员工晋升到资深级别）转移至 $(1,1,1)$，奖励为 $30-3=27$;

状态将以概率 $0.07 * 0.7$（即资深员工离职，初级员工没有晋升）转移至 $(0,1,0)$，奖励为 30-$3=27$。

状态将以概率 0 转移至 (0，0，0)，奖励为 30-$3=27$（其中 3 是雇用初级员工的成本，此处的奖励 30 来自使用人工智能工具工作的资深员工；）。

5.3.3 行动 (0,1,0)

如果行动为 (0，1，0)，资深员工不采用人工智能，而初级员工采用人工智能，且企业不进行招聘。

随后，状态将以 (0.07)*(0.3) 的概率转移到 (1,0,0)（其中 0.07 是资深员工离开企业的概率。资深员工离开，初级员工能力提升），并获得奖励 $(10+8-3)=15$。（在这种情况下，他们的奖励可以相加：资深员工在没有人工智能时获得 10，初级员工在使用人工智能时获得 5（扣除成本后））

状态将以 $0.93*0.7$ 的概率转移到 (1,1,0)（资深员工不离开，初级员工的能力也没有提升），并产生奖励 $10+5=15$;

状态将以 $0.93*0.3$ 的概率转移到 (1,1,1)（资深员工不离开，初级员工能力提升至资深员工水平），并产生奖励 $10+5=15$。

状态将以概率 $0.07*0.7$ 转移到 (0，1，0)（即资深员工离职，初级员工未得到提升），并产生奖励 $10+5=15$。

状态将以概率 0 转移到 (0，0，0)，奖励为 $10+5=15$。

5.3.4 行动 (1,0,0)

如果行动为 (1，0，0)，资深员工采用人工智能，但初级员工不采用人工智能；并且企业不进行招聘。

随后，状态将以 (0.07)*(0.35) 的概率转移到 (1，0，0)，其中 0.07 是资深员工离开企业的概率，奖励为 $(30-3)=27$。

状态将以概率 $0.93*0.65$ 转移到 (1，1，0)（资深员工未离职且初级员工未提升至资深级别），并产生奖励 $30-3=27$。

状态将以 $0.93*0.35$ 的概率转移至 (1，1，1)（资深员工不会离职，初级员工晋升至资深级别），并产生奖励 $30-3=27$。

状态将以 $0.07*0.65$ 的概率转移至 (0，1，0)（即资深员工离职，初级员工没有晋升），并产生奖励 30-$3=27$。

状态将以 0 的概率转移至 (0，0，0)，并产生奖励 30-$3=27$（其中 3 是雇用初级员工的成本，此处的奖励 30 来自资深员工使用人工智能工具工作）。

5.3.5 行动 (0,1,1)

如果行动为 (0，1，1)，资深员工不使用人工智能，初级员工采用人工智能；企业进行招聘。

随后，状态将以 (0.07)*(0.3)*0.09 的概率转移到 (1,0,0)（其中

0.07 是资深员工离开企业的概率），并获得 (10 + 5)-2=13 的奖励。

状态将以概率 $0.93*0.7+0.07*0.3*0.91$ 转移到 (1，1，0)（资深员工未离职且初级员工未提升；或资深员工离职、初级员工提升且招聘成功），奖励为 $10+5-2=13$。

状态将以概率 $0.93*0.3$ 转移到 (1，1，1)（资深员工未离职且初级员工提升至资深级别），奖励为 $10+5-2=13$。

状态将以概率 $0.07*0.7$ 转移到 (0，1，0)（即资深员工离职，初级员工未提升），奖励为 $10+5-2=13$。

状态将以概率 0 转移到 (0，0，0)，奖励为 $10+5-2=13$。

5.3.6 行动 (1,0,1)

如果行动为 (1,0,1)，资深员工使用人工智能，初级员工不采用人工智能，并且企业进行招聘。

随后，状态将以 (0.07)*(0.35)*0.09 的概率转移至 (1,0,0)（其中，0.07 是资深员工离开企业且招聘未成功的概率），并获得奖励 $30-3-2=25$。

状态将以 $0.93*0.65+0.07*0.35*0.91$ 的概率转移至 (1,1,0)（资深员工未离开且初级员工未改善；或资深员工离开、初级员工有所改善且招聘成功），并获得奖励 $30-3-2=25$。

状态将以 $0.93*0.35$ 的概率转移至 (1,1,1)（资深员工未离开且初级员工提升至资深级别），并获得奖励 $30-3-2=25$。

状态将以概率 $0.07*0.65$ 转移到 (0，1，0)（即资深员工离职，初级员工没有进步），并获得奖励 $30-3-2=25$。

状态将以概率 0 转移到 (0，0，0)，奖励为 $30-3-2=25$。

5.4 状态 (1,1,1)

假设当前状态为 (1，1，1)，即这家公司既有初级员工也有资深员工；并且在必要时，初级员工已准备好接替资深员工。

5.4.1 行动 (1,1,1)

如果行动为 (1，1，1)，资深员工使用人工智能，初级员工采用人工智能，且企业进行招聘。

然后，状态将以 (0.07)*0.09 的概率转移到 (1,0,0)（其中 0.07 是

资深员工离开企业且招聘未成功的概率），并获得奖励 $30-3-2=25$。

状态将以概率 $0.07*0.91$（资深员工离职且招聘成功）转移到 (1，1，0)，奖励为 $30-3-2=25$。

状态将以概率 0.93（资深员工未离职）转移到 (1，1，1)，奖励为 $30-3-2=25$。

状态将以概率 0（即资深员工离职，初级员工未提升）转移到 (0，1，0)，奖励为 $30-3-2=25$。

状态 将转移到 (0，0，0)，概率为 0，奖励为 $30-3-2=25$.

5.4.2 行动 (1,1,0)

如果行动为 (1,1,0)，资深员工使用人工智能，初级员工采用人工智能，且企业不进行招聘。

随后，状态将以 (0.07) 的概率转移到 $(1,0,0)$（其中 0.07 表示资深员工离开企业且初级员工晋升的概率），并获得奖励 $30-3=27$。

状 态将以 0 的概率转移到 $(1,1,0)$，并获得奖励 $30-3=27$。

状态将以 0.93 的概率转移到 $(1,1,1)$（资深员工不会离开），并获得奖励 $30-3=27$。

状态将以 0 的概率转移到 $(0,1,0)$（即资深员工离开，初级员工没有进步），并获得奖励 $30-3=27$。

状态将以概率 0 转移到 (0，0，0)，奖励为 $30-3=27$。

5.4.3 行动 (0,1,0)

如果行动为 (0，1，0)，资深员工不使用人工智能，初级员工采用人工智能，且企业不进行招聘。

随后，状态将以概率 (0.07) 转移到 (1，0，0)，奖励为 $10+5=15$。

状 态将以概率 0 转移到 (1，1，0)，奖励为 $10+5=15$。

状态将以 0.93 的概率转移到 $(1,1,1)$，奖励为 $10+5=15$,

状态将以 0 的概率转移到 $(0,1,0)$（即初级员工不会改进），奖励为 15。

状态将以概率 0 转移到 (0，0，0)，奖励为 15。

5.4.4 行动 (1,0,0)

如果行动为 (1，0，0)，资深员工使用人工智能，初级员工不采用人工智能，且企业不进行招聘。

然后，状态将以 (0.07) 的概率转移到 (1,0,0)（其中 0.07 是资深员工离开企业的概率），并获得奖励 $30-3=27$。

状态将以 0 的概率转移到 (1,1,0)，并获得奖励 $30-3=27$。

状态将以 0.93 的概率转移到 (1,1,1)（资深员工不会离开，并获得奖励 $30-3=27$。

状态将以 0 的概率转移到 (0,1,0)，奖励为 $30-3=27$。

状态将以 0 的概率转移到 (0,0,0)，并获得奖励 $30-3=27$。

5.4.5 行动 (0,1,1)

如果行动为 (0,1,1)，资深员工不使用人工智能，初级员工采用人工智能，且企业进行招聘。（请注意，在这种情况下，使用人工智能的初级员工与采用人工智能的资深员工相同）

然后，状态将以概率 (0.07)*0.09 转移到 (1,0,0)（其中 0.07 是资深员工离开企业且招聘未成功的概率），奖励为 $10+5-2=13$。

状态将以概率 $0.07*0.91$（或资深员工离开且招聘成功）转移到 (1,1,0)，奖励为 $10+5-2=13$。

状态将以概率 0.93（资深员工不离开）转移到 (1,1,1)，奖励为 $10+5-2=13$。

状态将以概率 0 转移到 (0，1，0)（即资深员工离职，初级员工没有进步），奖励为 $10+5-2=13$。

状　　态将以概率 0 转移到 (0，0，0)，奖励为 $10+5-2=13$.

5.4.6 行动 (1,0,1)

如果行动为 (1，0，1)，资深员工使用人工智能，初级员工不采用人工智能，且企业进行招聘。

那么　　状态将以 0.07*0.09, $30-3-2=25$ 的概率转移到 (1,0,0)。

状态将以 $0.07*0.91$ 的概率转移到 (1,1,0)（资深员工离职且招聘成功），并获得奖励 $30-3-2=25$。

状态将以 0.93 的概率转移到 (1,1,1)（资深员工不离职），奖励为 $30-3-2=25$。

状态将以 0 的概率转移到 (0,1,0)，奖励为 $30-3-2=25$。状态将以 0 的概率转移到 (0,0,0)，奖励为 $30-3-2=25$。

5.5 状态 (1,0,0)

假设当前状态为 (1，0，0)，也就是说，这家公司只有资深员工。

5.5.1 行动 (1,1,1)

如果采取行动 (1，1，1)，资深员工使用人工智能，初级员工采用人工智能，且企业进行招聘。

那么，状态将以 0.93*0.09 的概率转变为 (1，0， 0)（其中 0.93 是资深员工不离开企业且招聘未成功的概率），并获得奖励 $30-2=28$。

状态将以 $0.93*0.91$ 的概率转变为 (1， 1，0)（资深员工未离开且招聘成功），并获得奖励 $30-2=28$。

状　　态将以概率 0 转移至 (1，1，1)，奖励为 $30-2=28$。

状态将以概率 $0.07*0.91$（即资深员工离职且招聘成功）转移至 (0，1，0)，奖励为 $30-2=28$。

状态将以概率 $0.07*0.09$ 转移至 (0，0，0)，奖励为 $30-2=28$。

5.5.2 行动 (1,1,0)

如果行动为 (1，1，0)，资深员工使用人工智能，初级员工也采用人工智能（但没有初级员工），且企业不进行招聘。

随后，状态将以 0.93 的概率转移到 (1,0,0)（其中 0.07 是资深员工离开企业的概率），并获得 30 的奖励。

状态将以 0 的概率转移到 (1,1,0)，并获得 30 的奖励。

状态将以 0 的概率转移到 (1,1,1)（资深员工不会离开，且初级员工晋升至资深级别），并获得 30 的奖励。

状态将以 0 的概率转移到 (0,1,0)，并获得 30 的奖励。

状态将以概率 0，0，0 转移，获得奖励 30。

5.5.3 行动 (0,1,0)

如果行动为 (0，1，0)，资深员工不使用人工智能，而初级员工采用人工智能，且企业不进行招聘。

随后，状态将以概率 (0.93) 转移到 (1，0，0)，并获得奖励 10（资深员工不使用人工智能时的奖励为 10）。

状态将以概率 0 转移到 (1，1，0)，并获得奖励 10。

状态将以概率 0 转移到 (1，1，1)，并获得奖励 10。

状态将以概率 0 转移到 (0，1，0)（即初级员工没有进步），并获得奖励 10。

状态将以概率 0。07 转移到 (0，0，0)，并获得奖励 10。

5.5.4 行动 (1,0,0)

如果行动为 (1,0,0)，资深员工使用人工智能，初级员工不采用人工智能（没有初级员工），且企业不进行招聘。

随后，状态将以 (0.93) 的概率转移到 (1,0,0)（资深员工离开企业的概率为 0.07），奖励为 30。

状态将以 0 的概率转移到 (1,1,0)，奖励为 30。

状态将以 0 的概率转移到 (1,1,1)（资深员工不会离开，且初级员工晋升至资深级别），奖励为 30。

T状态将以 0 的概率转移到 (0,1,0)，奖励为 30。

状态将以 0,0,0) 为目标转移，概率为 0.07，奖励为 30。

5.5.5 行动 (0,1,1)

如果行动为 (0,1,1)，资深员工不使用人工智能，初级员工采用人工智能，并且企业进行招聘。

随后，状态将以 (1,0,0) 为目标转移，概率为 (0.93)*0.09，奖励为$10 - 2 = 8$。

状态将以 (1,1,0) 为目标转移，概率为 $0.93 * 0.91$，奖励为$10 - 2 = 8$。

状态将以 (1,1,1) 为目标转移，概率为 0（资深员工不离职），奖励为 $10 - 2 = 8$。

状态将以概率 $0.07 * 0.91$ 转移到 (0,1,0)（即资深员工离职且招聘成功），并获得奖励 $10 - 2 = 8$。

状态将以概率 $0.07*0.09$ 转移到 (0,0,0)，并获得奖励 $10-2=8$。

5.5.6 行动 (1,0,1)

如果行动为 (1，0，1)，资深员工使用人工智能，初级员工不采用人工智能，且企业进行招聘。

随后，状态将以概率 $0.93*0.09$ 转移至 (1，0，0)（其中 0.93 表示资深员工未离开企业且招聘未成功的概率），并获得奖励 $30-2=28$。

状态将以概率 $0.93*0.91$ 转移至 (1，1，0)（资深员工未离开且招聘成功），并获得奖励 $30-2=28$。

状态将以概率 0 转移至 (1，1，1)，并获得奖励 $30-2=28$。

状态将以概率 $0.07*0.91$ 转移至 (0，1，0)（即资深员工离开、招聘成功），并获得奖励 $30-2=28$。

状态 将以概率 $0.07*0.09$ 转移至 (0，0，0)，并获得奖励 $30-2=28$.

5.6 数值结果分析

为方便表示，我们按照前文所述，使用向量表示法表示每个状态。具体而言，S_0 =(0，0，0)、S_1 =(0，1，0)、S_2 =(1，0，0)、S_3 =(1，1，0)、S_4 =(1，1，1)。同样，我们使用向量表示法表示每个行动，其中 A_0 =(1，0，0)、A_1 =(1，1，0)、A_2 =(1，0，1)、A_3 =(0，1，1)、A_4 =(1，1，1)，以及 A_5 =(0，1，0)。

在图7中，我们绘制了每个状态下的最优动作。这里，行动 i 意味着行动 S_i。从图中可以看出，在状态 S_0 和 S_2下，最优选择是行动A_2；而在所有剩余状态下，最优选择是行动 A_0。状态 S_0 和S_2 表示企业当前没有初级员工的情况。在这些状态下，最优动作 A_2是资深员工采用人工智能，且企业雇用新的初级员工。尽管在没有初级员工的情况下，人工智能能够使资深员工独自创造较高的奖励，但最优动作仍然是雇用初级员工，并将他们作为储备劳动力，即使初级员工此时并未为提高奖励作出贡献。

相比之下，剩余状态代表企业中存在初级员工的情况。在这些状态下，偏好行动是 A_0：初级员工不使用AI卸载，资深员工采用人工智能，且企业不雇用额外员工。这使初级员工能够通过人力努力发展技能，而这些培训为他们未来成为资深员工提供了机会。如果初级员工使用卸载，其技能水平的提升可能会更加缓慢，因而无法及时成为资深员工并产生更高奖励。

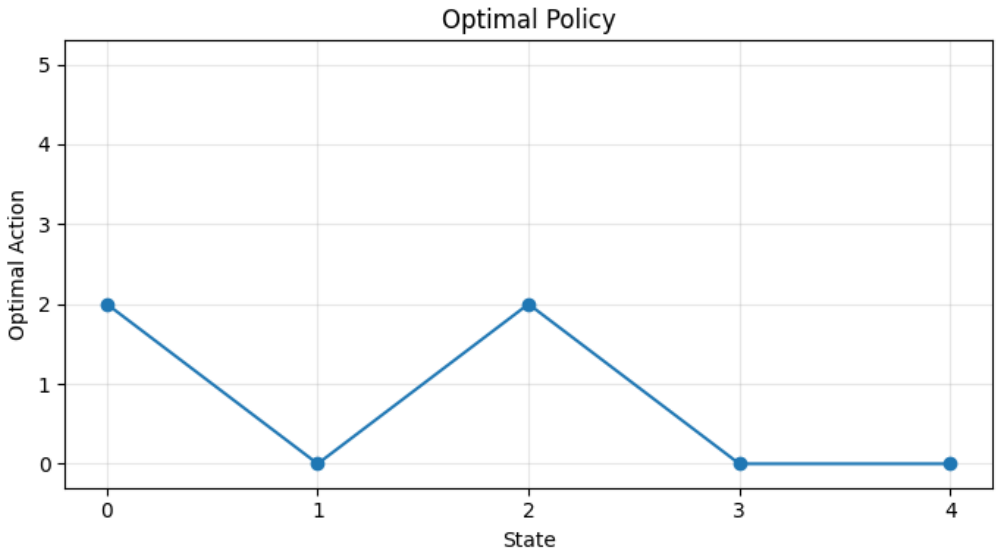

图7：最优动作与状态。此图展示了每个状态下的最优动作。此处，行动编号 i对应行动 A_i。

6 人工智能智能体经济学的马尔可夫决策过程模型

我们考虑 n 智能体之间的交易博弈，其中可能同时包括人类和人工智能智能体。

6.1 多智能体经济交换问题

我们考虑 n 个智能体，其中 n 是一个正整数。每个智能体既可以充当生产者，也可以充当消费者。请注意，在我们的模型中，“智能体”可以代表人类代理、AI智能体，甚至是人工智能与人类协作代理的实体（例如，一个国家的所有人可以在国际贸易中建模为一个智能体；一家科技公司拥有并控制的所有AI智能体可以建模为一个单一的“大型”智能体，提供产品/服务）。我们考虑一个支出矩阵 $P \in \mathbb{R}^{n\times n}$，其中 P 是一个马尔可夫链转移“概率”矩阵。我们将 P 中第 i行、第 j列的元素记为 P_{ij}。对于矩阵 P，对每个 $1 \le j \le n$都有 $\sum_{i=1}^{n} P_{ij} = 1$。数值 P_{ij} 表示第 j个参与者用于向第 i 个智能体购买产品/服务的货币比例。请注意，马尔可夫链转移概率矩阵中的数值并不代表其物理意义上的概率：它们仅表示一个智能体花费在向另一个智能体购买产品/服务上的资金比例。

我们用 $x^t \in \mathbb{R}^{n\times 1}$ 表示时间索引为 t 时的货币向量，其中 $t \ge 0$ 是一个整数。我们假设 x^t 的每个元素均为非负数，表示每个智能体在时间索引为 t 时所拥有的货币

。我们将连续时间索引之间的时间间隔称为一个回合。我们有一个效用矩阵 $U \in \mathbb{R}^{n\times n}$，其元素均为非负数。我们将 i 的第 i 行第 j 列元素记为 U_{ij}，该元素表示：如果第 j 个智能体花费一美元从第 i 个智能体处购买产品 / 服务，则第 i 个参与者为第 j 个智能体提供的效用。

每个智能体的目标都是最大化其获得的效用。表面上看，为了最大化第 j 个智能体的效用，第 j 个智能体应将其所有货币都花在每花一美元就能为其提供最高效用的智能体身上，即具有最大 U_{ij} 的智能体 i 。然而，出乎意料的是，我们认为事实并非如此，最优的支出策略应取决于生产者与消费者网络的结构。我们注意到，在时间索引 $(t+1)$，货币向量更新为

$$x^{t+1} = Px^t.$$

我们也可以令支出矩阵 P 随时间 t 变化，并将其记为 P^t。支出矩阵 P^t 描述了所有智能体的动作：其第 j 列是第 j 个智能体在时间索引 t 的动作。智能体 i 的状态由该智能体拥有的财富来描述。

下面，我们举一个例子，说明多智能体MDP模型为如何在经济博弈中最佳地整合人工智能智能体提供了解决方案。

6.2 经济交易博弈示例

我们介绍一个经济博弈的示例。假设有 $n = 3$ 个智能体：其中两个是人类智能体 H_1 和 H_2；另一个是 AI 智能体。

假设系统中只有一个单位的总财富。假设 $\gamma < 1$是折扣因子。当拥有一个单位的财富时，智能体 H_1 有两种可能的行动。第一种行动是向AI智能体购买服务，此时 H_1 将一个单位的财富给予AI智能体，并获得100的环境奖励（即效用）；第二种行动是向 H_2购买服务，此时 H_1 将一个单位的财富给予H_2 ，并获得5的环境奖励。当拥有一个单位的财富时，智能体 H_2 有两种可能的行动。第一种行动是向AI智能体购买服务，此时 H_2 将一个单位的财富给予AI智能体，并获得80的环境奖励；第二种行动是向 H_1购买服务，此时 H_2 将一个单位的财富给予 H_1，智能体 H_2 获得100的环境奖励。当拥有一个单位的财富时，AI智能体有三种行动：第一种行动是亲自完成工作，获得1000的环境奖励；第二种行动是以一个单位财富的成本向 H_1 购买服务，获得30的环境奖励；第三种行动是以一个单位财富的成本向 H_2 购买服务，获得30的环境奖励。当智能体没有任何财富时，无法向他人购买服务，也无法亲自完成工作。

那么，三个智能体的以下策略构成一个均衡：智能体 H_1 只要拥有一个单位的财富，就总是向智能体 H_2 购买服务；智能体 H_2 只要拥有一个单位的财富，就总是向智能体 H_1 购买服务；AI智能体只要拥有一个单位的财富，就总是向自身购买服务。

为证明这是一个均衡，我们需要说明：在其他人保持其策略不变的情况下，任何人都无法通过改变自己的策略来提高其最优值。首先，我们注意到，对于 AI智能体 而言，购买自己的服务始终是最优的，因为在每个时间步，它都能获得可能的最大奖励 1000。假设智能体 H_1 的最优策略是在拥有一单位财富时购买 AI智能体 的服务。那么，它只能获得 100 的折扣奖励，因为随后 AI智能体 会将这笔财富用于购买自己的服务。然而，如果智能体 H_1 转而将这一单位财富用于购买智能体 H_2 的服务，则获得的折扣奖励将为 $\frac{5}{1-\gamma^{20}}$。因此，当 γ 接近 1 时，或者我们考虑未折扣奖励时，这一奖励将大于通过 H_1 从 AI智能体 处购买服务的策略所获得的奖励。对于智能体 H_2 也可以得出同样的结论。因此，这确实是一个纳什均衡。我们注意到，在这一均衡中，人类智能体的最优策略并不是向 AI智能体 购买服务，尽管 AI智能体 提供了更高的即时奖励。

对于上述均衡，人们或许会疑惑：为什么人类代理的最优策略不是向最强大的 AI 智能体购买产品，因为它每花一美元都能提供最高效用？这是因为，在我们的模型中，智能体的支出模式或策略不仅会影响其消费所获得的即时效用，还会影响其所处的财富状态。

7 结论与未来方向

本文提出了一个基于多智能体马尔可夫决策过程（MDP）的通用强化学习（RL）框架，用于确定和分析人工智能技术的最优应用。该框架能够对人工智能应用的长期动态影响进行建模，并据此找到最优人工智能采用策略。在我们的建模框架中，人工智能应用的负面影响通过 MDP 中的状态转移或负奖励来建模。所提出的多智能体强化学习框架同时纳入了人类智能体和人工智能智能体。

我们模型的核心思想是，纳入人工智能技术和人工智能智能体，不仅可能增加人工智能智能体的即时奖励，还可能改变相关人类智能体和人工智能智能体的状态。在经济中，采用人工智能可能影响智能体之间的就业机会和财富分配，从而改变智能体所处的财富状态或就业机会状态。对于企业而言，使用人工智能技术取代初级员工，在面临资深员工流失或新颖的未来挑战时，可能导致劳动力准备不足。

未来　　工作中，探索更通用的策略非常重要，例如依赖历史的　　t

在采用人工智能技术方面的多智能体强化学习策略，Lu 等人（2026年）。

参考文献

Abrardi, L.、C. Cambini 和 L. Rondi。2022年。人工智能、企业与消费者行为：一项综述。《经济调查期刊》36：969–91。

Agrawal, A. K. 等。2023年。通用技术采用的相似性与差异。美国国家经济研究局。

Akin, M. S. 2026年。人工智能何时会增加劳动需求？可见于 *SSRN 6849279*。

Aouad, A.、T. Lykouris 和 H. Zhong。2026年。人类与人工智能生产率悖论：对技能、努力与人工智能辅助之间相互作用的建模。*arXiv* 预印本 *arXiv:2605.11350*。

Bauer, A. 2026年。作为信号的备用方案：在人工智能不断改进的信任品市场中，保留的人类技能、责任与能力信号。*arXiv* 预印本 *arXiv:2608.04276*。

Bondi, T., and G. Johnson. 2026. 技能退化与人工智能生产率测量。*AvailableatSSRN6169671* 。

Cai, H., Z. Lu, C. Xu, W. Xu, and J. Zheng. 2026. 人工超级智能或许毫无用处：多个人工智能智能体经济中的均衡。*arXiv*预印本 *arXiv:2603.00858*。

Cai, H., C. Xu, and W. Xu. 2023. 要不要使用人工智能，要不要购买本地产品：实际价格的数学理论。*arXiv*预印本 *arXiv:2305.05134* 。

Caosun, M., and S. Aral. 2026. 增强陷阱：人工智能生产率与认知卸载的成本。*arXiv*预印本 *arXiv:2604.03501* 。

Dasaklis, T. K., P. G. Giannopoulos, D. Koutras, V. Malamas 和 P. T. Chountalas。2026年。《人力资源管理中的大语言模型：应用、开放问题与未来研究方向的系统文献回顾》

献回顾。《技术预测与社会变革》231:124800–。

Falk, B. H. 和 G. Tsoukalas。2026年。人工智能裁员陷阱。*arXiv* 预印本 *arXiv:2603.20617*。

Goldfarb, A. 和 C. Tucker。2019年。数字经济学。《经济文献杂志》57：3–43。

Hammond, L. 等。2025年。先进人工智能带来的多智能体风险。*arXiv* 预印本 *arXiv:2502.14143* 。

Kapoor, S. 等。2024年。具有重要意义的人工智能智能体。*arXiv* 预印本 *arXiv:2407.01502* 。

Kimura, T. 2026年。效率还是分配：人工智能裁员陷阱中的隐性权衡。可在 *SSRN 6936479* 获取 。

Li, J. 等。2024年。更多智能体就是你所需要的一切。*arXiv* 预印本 *arXiv:2402.05120* 。

Lu, Z.、A. R. Mudireddy、S. Alqahtani 和 W. Xu。2026年。历史至关重要：采用异构多智能体强化学习的元策略委派。