

计算机

发布时间: 2023-03-23

证券研究报告 / 行业深度报告

GPT 将如何影响我们的工作?

优于大势

上次评级: 优于大势

报告摘要:

OpenAI 官网发布了最新研究论文 *GPTs are GPTs: An early look at the labor market impact potential of large language models*, 论文中对 LLM 语言模型和 GPT 对美国不同职业的潜在影响进行了探讨。论文研究结果表明, 在美国:

- 多数职业将受到 GPT 的冲击:** 80%的工人有至少 10%的任务可以被 GPT 减少 $\geq 50\%$ 的工作时间; 19%的工人有至少 50%的任务可以被 GPT 减少 $\geq 50\%$ 的工作时间;
- GPT 的影响横跨各类薪资层级:** 尽管存在部分特殊情况, 但整体来看, 工资越高, 受 GPT 冲击的程度越大;
- 职业技能与 GPT 的冲击程度有关:** 科学和批判性思维技能最不容易受 GPT 冲击, 而编程和写作技能受影响的程度最高;
- 高学历更容易受到 GPT 的冲击:** 持有学士、硕士和更高学位的人比没有正规教育学历的人更容易受到 GPT 的冲击;
- 在职培训时间时长与 GPT 冲击程度呈负相关:** 在职培训时长最长的职业收入水平偏低, 且受 GPT 冲击程度最低, 而没有在职培训或只需实习的工作则表现出更高的收入水平和更容易受 GPT 冲击的属性。
- 证券相关和数据处理行业受 GPT 影响程度最高:** 在人类打分和 GPT 打分模式下, 证券商品合约及其他金融投资和数据处理托管分别是受 GPT 冲击程度最高的行业; 在进一步开发 GPT 衍生功能的情况下, 量化分析师的所有任务将被 GPT 降低 50%以上的工作时间。

此外, 我们运用论文中的方法, 对中国证券行业情况进行了类似的打分和统计, 并得出如下结论:

- 受 GPT 冲击的程度:** 二级卖方分析师 > 投行一级市场 > 基金经理。
- 在经过专业知识训练的 LLM 和 GPT 的帮助下, 卖方分析师可能有 82%的任务将被减少 50%以上的工作时间, 基金经理可能有 55%的任务被减少 50%以上的工作时间。

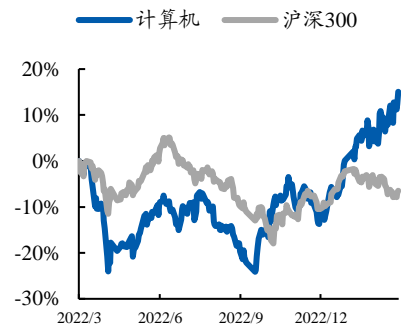
与美国的对比: 国内卖方分析师、一级市场(投行)和基金经理受到 GPT 冲击的程度可能略好于美国证券行业受冲击的平均水平, 这可能与量化分析师比例和资本市场有效性的差异有关。

风险提示: AI 技术创新不及预期, 行业竞争加剧

重点公司主要财务数据

重点公司	现价	EPS			PE			评级
		2021	2022(E)	2023E	2021	2022(E)	2023E	
金山办公	360.82	2.26	2.42	2.97	117.32	109.29	121.57	买入
恒生电子	45.93	1.01	0.57	0.97	61.53	80.29	47.47	买入
科大讯飞	56.71	0.70	0.25	0.77	75.01	228.33	74.03	买入

历史收益率曲线



涨跌幅 (%)	1M	3M	12M
绝对收益	8%	33%	12%
相对收益	10%	29%	18%

行业数据

成分股数量 (只)	346
总市值 (亿)	44935
流通市值 (亿)	23866
市盈率 (倍)	118.5
市净率 (倍)	6.0
成分股总营收 (亿)	3520
成分股总净利润 (亿)	243
成分股资产负债率 (%)	36.1

相关报告

《从海外映射看 ChatGPT 在 A 股的投资机会》

--20230206

证券分析师: 黄净

执业证书编号: S0550522010001

18680586451 huangjing@nesc.cn

研究助理: 吴雨萌

执业证书编号: S0550122040013

18901997197 wuyum@nesc.cn

目 录

1.	总结：GPT 对工作的冲击将跨越各个职业.....	4
2.	统计指标来源及解释	4
2.1.	数据来源	4
2.1.1.	美国职业、工作活动和任务数据的来源.....	4
2.1.2.	工资、就业及人口数据来源.....	5
2.2.	暴露度 Exposure: 用于衡量 GPT 对各职业的冲击程度.....	5
3.	研究结论：30%的职业或任务将受到 GPT 冲击.....	7
4.	研究结论：工资水平与 GPT 冲击程度呈正相关.....	8
5.	研究结论：科学和批判性思维是受 GPT 冲击最小的技能	10
6.	研究结论：学历水平和在职培训时长与 GPT 冲击程度相关	12
7.	研究结论：证券投资和数据处理可能是受冲击程度最高的职业.....	14
8.	对国内的探讨：卖方分析师≥80%的工作可能受 GPT 冲击	20
	风险提示.....	23

图表目录

图 1:	经济体暴露度（左图：职业和任务暴露度的分布，右图：工人和任务暴露度的分布）	8
图 2:	各职业就业水平的暴露度分布（人类打分和 GPT 打分模式）	8
图 3:	各职业工资的暴露度分布（人类打分和 GPT 打分模式）	9
图 4:	五个工作区对应的职业暴露度 β	12
图 5:	不同行业受 GPT 的影响程度（人类打分模式）	16
图 6:	不同行业受 GPT 的影响程度（GPT 打分模式）	17
图 7:	全要素生产力和劳动生产力对暴露度并无显著影响	18
表 1:	O*NET 数据库中职业、任务和 DWA 的示例.....	5
表 2:	GPT 和人类打分两种方式的一致性和皮尔逊相关系数情况	6
表 3:	统计数据汇总（人类打分和 GPT 打分）	7
表 4:	O*NET 技能和暴露度的 OLS 回归结果	10
表 5:	O*NET 技能的列表和定义.....	11
表 6:	不同工作区的分类方法及对应的暴露度	13
表 7:	按准入学学历分职业暴露度的平均值	13
表 8:	按在职培训时长分职业暴露度的平均值	13
表 9:	不同 GPT 技术水平下，受冲击程度最高的职业	15

表 10: 对 GPT 没有任何暴露度的行业主要以体力劳动为主	19
表 11: 二级卖方分析师、一级市场投行和基金经理受 GPT 冲击的情况	20
表 12: 二级卖方分析师、一级市场投行和基金经理的任务列表	21
表 13: 论文中关于暴露度分类的详细定义	22

1. 总结：GPT 对工作的冲击将跨越各个职业

3月17日，OpenAI 官网发布最新研究论文 GPTs are GPTs: An early look at the labor market impact potential of large language models，对 LLM 语言模型，特别是 GPT，对美国不同职业和行业的潜在影响进行了探讨。我们将论文中的结论进行了汇总：

- 1、**多数职业将受到 GPT 的冲击**：80%的工人有至少 10%的任务可以被 GPT 减少 $\geq 50\%$ 的工作时间；19%的工人有至少 50%的任务可以被 GPT 减少 $\geq 50\%$ 的工作时间；
- 2、**GPT 的影响横跨各类薪资层级**：尽管存在部分特殊情况，但整体来看，工资越高，受 GPT 冲击的程度越大；
- 3、**职业技能与 GPT 的冲击程度有关**：科学和批判性思维技能最不容易受 GPT 冲击，而编程和写作技能受影响的程度最高；
- 4、**高学历更容易受到 GPT 的冲击**：持有学士、硕士和更高学位的人比没有正规教育学历的人更容易受到 GPT 的冲击；
- 5、**在职培训时间时长与 GPT 冲击程度有关**：在职培训时长最长的职业收入水平偏低，且受 GPT 冲击程度最低，而没有在职培训或只需实习的工作则表现出更高的收入水平和更容易受 GPT 冲击的属性。
- 6、**证券相关和数据处理行业受 GPT 影响程度最高**：在人类打分和 GPT 打分模式下，证券商品合约及其他金融投资和数据处理托管分别是受 GPT 冲击程度最高的行业；在直接调用 GPT 模型的情况下，口译笔译和数学家分别是受影响最大的职业；在进一步开发 GPT 衍生功能的情况下，数学家和会计审计则分别为受影响最大的职业。

2. 统计指标来源及解释

2.1. 数据来源

2.1.1. 美国职业、工作活动和任务数据的来源

论文中使用了 O*NET 27.2 数据库，包含 1016 种职业，以及各个职业的工作活动（Detailed Work Activities，简称 DWA）和任务（Task）。论文中对工作活动和任务给出了定义：

- **详细工作活动 DWA** 是由完成任务构成的综合操作，大多数工作活动与一个或多个任务相对应，该数据集中包括 2087 种 DWA；
- **任务 Task** 是某个特定职业的基础单位，一项任务可以与 0 个、1 个或多个 DWA 关联，且每个任务都有与之对应的职业，该数据集中包括 19265 种任务。

例如，对于职业“急症护理护士”，其工作活动 DWA 包括“操作诊断或治疗性医疗仪器或设备”和“准备医疗用品或设备”，其任务包括“设置、操作或监测侵入性设备和装置，例如结肠造口术或气管切开术设备、机械呼吸机、导管、胃肠道管和中心插管”。

表 1: O*NET 数据库中职业、任务和 DWA 的示例

Task ID	Occupation Title	DWAs	Task Description
14675	Computer Systems Engineers/Architects	Monitor computer system performance to ensure proper operation.	Monitor system operation to detect potential problems.
18310	Acute Care Nurses	Operate diagnostic or therapeutic medical instruments or equipment. Prepare medical supplies or equipment for use.	Set up, operate, or monitor invasive equipment and devices, such as colostomy or tracheotomy equipment, mechanical ventilators, catheters, gastrointestinal tubes, and central lines.
4668.0	Gambling Cage Workers	Execute sales or other financial transactions.	Cash checks and process credit card advances for patrons.
15709	Online Merchants	Execute sales or other financial transactions.	Deliver e-mail confirmation of completed transactions and shipment.
6529	Kindergarten Teachers, Except Special Education	-	Involve parent volunteers and older students in children's activities to facilitate involvement in focused, complex play.
6568	Elementary School Teachers, Except Special Education	-	Involve parent volunteers and older students in children's activities to facilitate involvement in focused, complex play.

数据来源：东北证券、GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

2.1.2. 工资、就业及人口数据来源

论文选取了美国劳工统计局 (Bureau of Labor Statistics, 以下简称 BLS) 提供的 2020 年和 2021 年职业就业系列中的就业和工资数据。该数据集包括职业名称、每个职业的工人数量、2031 年职业水平的就业预测、职业准入的教育水平以及获得职业能力所需的在职培训情况。BLS 数据库可以同 O*NET 数据库进行联动：通过当前人口调查 (Current Population Survey, 简称 CPS)，将 O*NET 中的任务和工作活动数据集与 BLS 劳动力人口统计数据联系起来，形成截面数据。

2.2. 暴露度 Exposure: 用于衡量 GPT 对各职业的冲击程度

论文中设定了**暴露度 Exposure**这一指标，作为重点讨论的对象。暴露度 Exposure 用于衡量 GPT 对特定工作活动和任务的冲击程度，即保证一项工作活动和任务完成质量相同的情况下，使用 GPT 或 GPT 驱动的系统是否能够将执行工作活动或任务的所需时间减少 50% 以上。

论文采用了两种暴露度的注释方式，分别为人工评分法与 GPT-4 评级法：

- **人工评分：**通过对 O*NET 数据库中的每一个工作活动 DWA 和任务进行人为归类并注释打分。
- **GPT-4 评级：**采用早期版本的 GPT-4 对工作活动和任务进行注释打分。

论文将暴露度分为以下三类：

E0 无暴露度：如果经验丰富的工人在高质量完成任务时所需的时间没有明显减少 50%，或使用 GPT 相关技术会降低工作活动/任务的完成质量，则定义为 E0 (例：需要高强度人际互动的任务)。

E1 直接暴露：在保证完成质量相同的前提下，如果通过 ChatGPT 或 OpenAI 直接访问 LLM 或 GPT-4 可以将完成工作活动或任务所需时间减少 50% 及以上，则将其定义为 E1 (例：指令编写、转换文本和代码的任务)。

E2 LLM+暴露：直接访问 LLM 不能将完成任务所需的时间减 > 50%，但在 LLM 基础上开发额外功能后可以达成目的，则定义该类工作活动和任务为 E2 (例：总结超

过 2000 字的文档并回答关于文档的问题)。

为了更为准确地衡量暴露度这一指标的统计学意义，论文中构建了三个度量指标， α 、 β 和 ζ ，分别衡量低、中、高水平下的 GPT 对各职业的冲击程度。其中， $\alpha=E1$ ，代表一个职业受 GPT 冲击程度的下限； $\beta=E1+0.5*E2$ ，其中 E2 的 0.5 倍权重旨在解释通过补充工具或应用程序来完成任务/工作活动需要额外计算的暴露度； $\zeta=E1+E2$ ，代表一个职业受 GPT 冲击程度的上限，可用于评估一项工作/任务对于 GPT 及 GPT 驱动的系统最大暴露度(即 GPT 进一步开发后，一项工作/任务受到的最大影响)。

表 2: GPT 和人类打分两种方式的一致性和皮尔逊相关系数情况

Comparison	γ	Weighting	Agreement	Pearson's
GPT-4, Rubric 1; Human	α	E1	80.8%	0.223
	β	E1 + .5*E2	65.6%	0.591
	ζ	E1 + E2	82.1%	0.654
GPT-4, Rubric 2; Human	α	E1	81.8%	0.221
	β	E1 + .5*E2	65.6%	0.538
	ζ	E1 + E2	79.5%	0.589
GPT-4, Rubric 1; GPT-4, Rubric 2	α	E1	91.1%	0.611
	β	E1 + .5*E2	76.0%	0.705
	ζ	E1 + E2	82.4%	0.680

数据来源：东北证券、GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

注：作者采用了两种 GPT-4 的打分规则，论文中作者应用了 GPT-4 规则 1 进行统计解释。

3. 研究结论：30%的职业或任务将受到 GPT 冲击

前文将暴露度 Exposure 这一指标的定义进行了描述，论文中还将暴露度的衡量指标 α 、 β 和 ζ 进行了统计数据的汇总。不论采取人类打分的方式还是 GPT-4 打分，暴露度 α 的均值在 0.14 左右，表示了从平均意义上说，15%左右的职业/任务暴露于 GPT，即 15%左右的工作可能会被现有的 LLM/GPT-4 降低 50%以上的工作时间。类似地，暴露度 β 和 ζ 均值分别在 0.3 和 0.5 左右，代表 30%/50%的职业或任务将受到中/高水平的 GPT 冲击，即减少工作时间 50%及以上。

表 3: 统计数据汇总（人类打分和 GPT 打分）

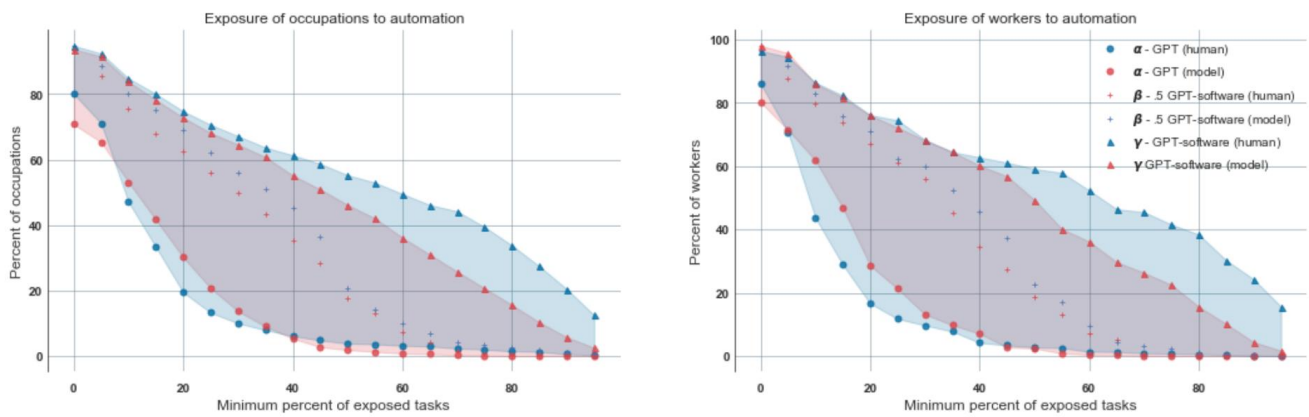
Occupation Level Exposure				
	Human		GPT-4	
	mean	std	mean	std
α	0.14	0.14	0.14	0.16
β	0.30	0.21	0.34	0.22
ζ	0.46	0.30	0.55	0.34
Task Level Exposure				
	Human		GPT-4	
	mean	std	mean	std
α	0.15	0.36	0.14	0.35
β	0.31	0.37	0.35	0.35
ζ	0.47	0.50	0.56	0.50

数据来源：东北证券、GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

4. 研究结论: 工资水平与 GPT 冲击程度呈正相关

论文探索了职业、工人分布程度与暴露度之间的关系。对于中等水平的 GPT (β) 来说, 约 19% 的工人有 50% 以上的任务将受到 GPT 的冲击, 80% 的工人有 10% 以上的任务受到了 GPT 的冲击; 18% 的职业中有 50% 以上的任务受到了 GPT 的冲击。从图表上看, 职业/工人百分比的暴露度分布相似, 表明 GPT 的冲击程度与不同职业下工人的情况无直接相关性。

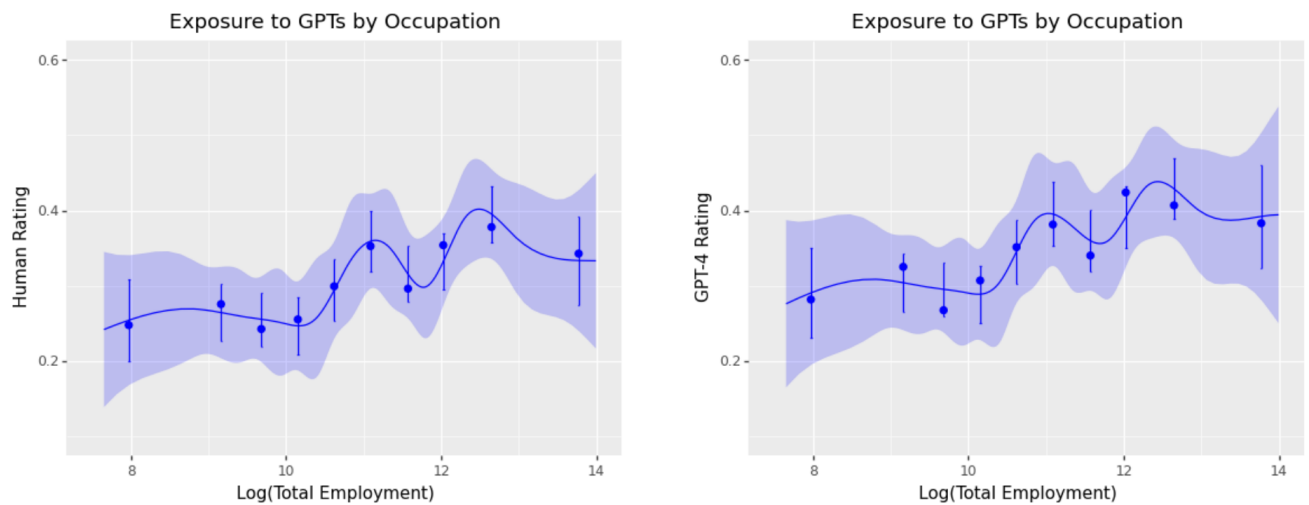
图 1: 经济体暴露度 (左图: 职业和任务暴露度的分布, 右图: 工人和任务暴露度的分布)



数据来源: 东北证券、GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

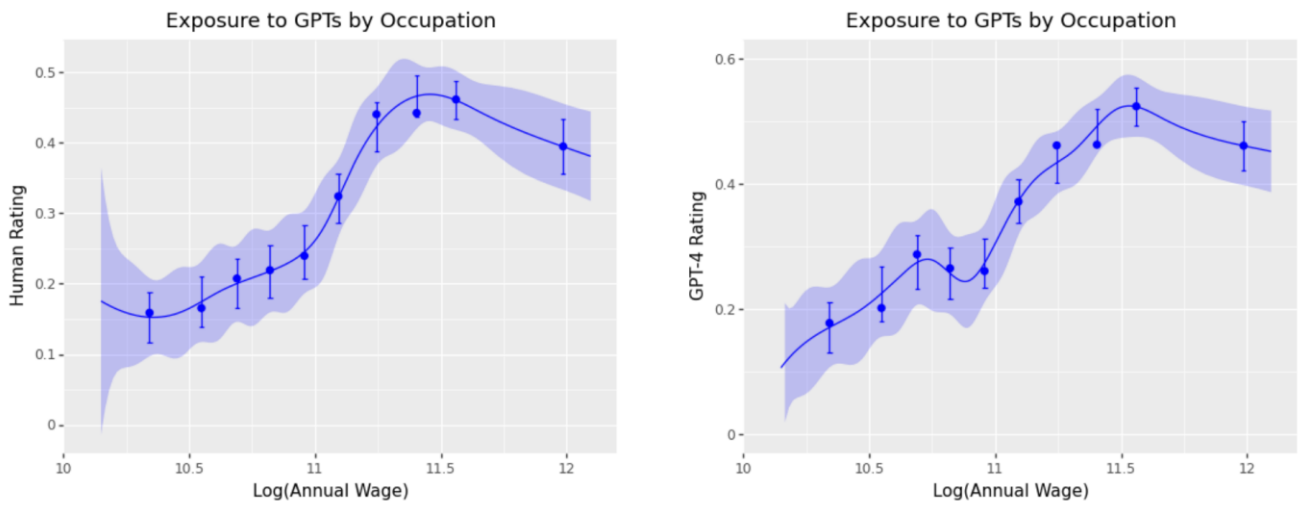
论文还对工资、就业水平与暴露度的相关性进行了探讨。两种打分模式下, 尽管存在一些高暴露度的低工资职业和低暴露度的高工资职业, 整体图表显示, **工资越高, 受 GPT 影响的程度也随之增加**。而 GPT 冲击程度与就业水平则并无显著关联。

图 2: 各职业就业水平的暴露度分布 (人类打分和 GPT 打分模式)



数据来源: 东北证券、GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

图 3: 各职业工资的暴露度分布 (人类打分和 GPT 打分模式)



数据来源: 东北证券、GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

5. 研究结论：科学和批判性思维是受 GPT 冲击最小的技能

论文研究了不同职业中技能重要性与 GPT 暴露度之间的关系。作者将 O*NET 数据库中的基本技能进行标准化，并将其与暴露度指标 (α , β , ζ) 进行回归分析，检验技能重要性和暴露度之间的关联度。结果表明，**科学和批判性思维技能 (Science and Critical Thinking)** 与暴露度呈强烈的负相关 (以 β 作为研究，相关系数分别 -0.23 和 -0.19)，即需要该技能的职业或任务不太可能受到 GPT 的冲击；相反，**编程和写作技能 (Programming and Writing)** 与暴露度呈现出强正相关 (相关系数分别为 0.62 和 0.47)，即涉及该技能的职业更容易受到 GPT 的冲击。

表 4: O*NET 技能和暴露度的 OLS 回归结果

	α (std err)	β (std err)	ζ (std err)
Constant	0.082*** (0.011)	-0.112*** (0.011)	0.300*** (0.057)
Active Listening	0.128** (0.047)	0.214*** (0.043)	0.449*** (0.027)
Mathematics	-0.127*** (0.026)	0.161*** (0.021)	0.787*** (0.049)
Reading Comprehension	0.153*** (0.041)	0.470*** (0.037)	-0.346*** (0.017)
Science	-0.114*** (0.014)	-0.230*** (0.012)	-0.346*** (0.017)
Speaking	-0.028 (0.039)	0.133*** (0.033)	0.294*** (0.042)
Writing	0.368*** (0.042)	0.467*** (0.037)	0.566*** (0.047)
Active Learning	-0.157*** (0.027)	-0.065** (0.024)	0.028 (0.032)
Critical Thinking	-0.264*** (0.036)	-0.196*** (0.033)	-0.129** (0.042)
Learning Strategies	-0.072* (0.028)	-0.209*** (0.025)	-0.346*** (0.034)
Monitoring	-0.067** (0.023)	-0.149*** (0.020)	-0.232*** (0.026)
Programming	0.637*** (0.030)	0.623*** (0.022)	0.609*** (0.024)

数据来源：东北证券、GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

欢迎访问：电子书学习和下载网站 (<https://www.shgis.com>)

文档名称：2023年GPT将如何影响我们的工作? .pdf

请登录 <https://shgis.com/post/1713.html> 下载完整文档。

手机端请扫码查看：

