

BAAI

智源社区 2021 AI 大调查 情况报告

2022 年 1 月

目录

前言	6
摘要	7
(一) 总体结论	7
(二) 学术界调研结论	9
(三) 工业界调研结论	10
(四) 投资界调研结论	11
一、受访者总体情况	12
1. 年龄分布	12
2. 性别分布	12
3. 地区分布	13
4. 最高学历分布	14
5. 所在领域分布	14
6. AI 研究方向	15
7. 编程语言	15
8. 常用的 AI 开发工具/框架/平台	16
9. 常用的机器学习算法	17
10. 常用的深度学习算法	17
11. 计算机视觉领域常用架构	18

12.自然语言处理领域常用架构	19
13.感兴趣的预训练研究方向	20
14.2021 年十大 AI 热点	21
15.学习 AI 知识与技能的渠道	22
16.获得 AI 信息的渠道	23
17.希望从 AI 社区中获得的资源	23
18.在 AI 社区中关注的功能	24
19.需要的支持资源	25
二、学术界受访者情况.....	26
1.学术界身份	26
2.科研团队规模	26
3.每周科研工作时长	27
4.每周教学工作时长	27
5.阻碍做出好科研的因素	28
6.学术界受访者的研究方向.....	28
7.学术界常用的深度学习算法.....	29
8.学术界感兴趣的预训练方向.....	30
9.学术界需要获得的支持	31
三、工业界受访者情况.....	33
1.工业界职位	33
2.AI 研发年限	33

3.月薪资范围	34
4.周工作时长	34
5.所在行业	35
6.工作单位性质	36
7.研发团队规模	37
8.企业融资轮次	37
9.AI 在企业中的使用情况.....	38
10.工业界受访者的研究方向	39
11.获得 AI 信息的渠道	40
12.希望从 AI 社区中获得的资源	40
四、投资界受访者情况.....	42
1.创业选择	42
2.投资选择	42
五、研究方法	44
(一) 问卷题目	44
(二) 数据清洗和分析方法.....	46
(三) 调研存在的局限性.....	47
免责声明	48

感谢以下成员对大调查工作的支持：

刘江对报告结构、内容等提供了方向性指导。

戴一鸣、李梦佳策划智源社区 2021 AI 大调查。

戴一鸣、李梦佳、钱曙光、吴阳参与制定问卷题目。

戴一鸣、钱曙光、常政、赵尚参与问卷发放和推广。

戴一鸣负责数据清洗、分析、可视化，并撰写报告。

智源研究院大模型研究中心、战略研究中心等成员，以及智源社区成员参与问卷前期测试，并提出反馈意见和建议。

前言

智源社区成立于 2019 年，现已成长为拥有 10 万人工智能领域用户，紧密联系 3000 多位 AI 顶级学者的庞大生态。为了了解广大成员当前的发展情况，探究受访者在科研探索、技能提升、职业发展等方面的需求，推动为成员提供更有价值和针对性的服务与资源，智源社区于 2021 年 11 月启动大调查活动，面向社区成员及广大 AI 领域的研究者和从业者，收集广大用户的意见。目前调查已截止，我们通过数据清洗、筛选、统计等方式进行分析，最终形成本报告，现将结果公布如下。

摘要

本次调查发现以下情况：

（一）总体结论

1. 受访者总体上较为年轻，绝大多数拥有本科及以上学历，近半数拥有研究生（含硕士、博士）学历。
2. 性别分布上，男性受访者数量是女性的两倍以上。
3. 受访者主要来自学术界和工业界，其中学术界受访者占比近半，工业界受访者占比超过三成。投资界受访者仅为学术界受访者的三分之一，工业界受访者的一半。
4. 受访者主要从事的研究方向包括计算机视觉、自然语言处理和机器学习等。
5. Python 是当前受访者中最常用的编程语言，随后是 C++、Java 和 SQL。
6. PyTorch 是目前受访者中最常用的 AI 开发工具，其用户数大约是 TensorFlow 用户数量的 1.5 倍。
7. 受访者最常用的机器学习算法是随机树或随机森林，其次是梯度提升机（XGboost）。
8. 受访者最常用的深度学习算法包括卷积神经网络（CNN）、循环神经网络（如 RNN、LSTM、GRU 等），Transformer 在近来开始崭露头角。

9.当前在预训练领域，受访者最关注的是多模态数据学习，其次是知识、常识和模型的融合。

10.根据投票情况，受访者选择的 2021 年人工智能领域十大热点事件如下：

- 1) David Card, Joshua D.Angrist 和 Guido W.Imbens 因对因果关系分析方法论的贡献获得诺贝尔经济学奖；
- 2) OpenAI 发布多模态预训练模型 DALL·E 和 CLIP；
- 3) 谷歌发布 1.6 万亿参数模型 Switch Transformer；
- 4) 智源研究院发布万亿规模超大规模智能模型悟道 2.0；
- 5) AlphaFold2/RosTTAFold 算法和相关数据集等开源；
- 6) 微软亚洲研究院提出 Swin Transformer 视觉预训练模型；
- 7) 何恺明团队提出 Masked Autoencoders 视觉预训练方法；
- 8) 清华、智源等提出 Pre-Train Prompt Tuning (PPT) 框架；
- 9) 斯坦福大学研究者利用 AI 和脑机接口实现“意念写字”；
- 10) 谷歌发布改进版视觉预训练模型 ViT。

在社区用户票选的十大进展中，有 5 项与预训练模型非常相关。预训练是 2021 年最受关注的领域之一。

11.受访者主要获得 AI 知识和技能的渠道是通过大学教育和工作经验，但自学和培训班已成为重要的补充渠道。

12.受访者主要获取 AI 信息渠道的方式是 GitHub，但也有许多受访者通过智源社区、知乎、公众号等方式获取信息。

13. “学习资源”是社区用户最希望获得的资源，包括学术活动、论文解读、知识课程等；超过半数的受访者主要希望获得学术活动资源。

14. 在 AI 社区中，近四成的用户关注的功能是阅读文章和讨论，以及创作文章、博客等。

15. 受访者总体上需要的支持包括专业交流和人脉圈，以及行业信息资讯等。

(二) 学术界调研结论

1. 学术界受访者中，学生数量是有教职研究者数量的 5 倍左右。

2. 学术界中，科研团队规模不超过 9 人的受访者近半。

3. 六成受访者每周的科研时长超过 20 小时。有一成的科研人员每周科研时间超过了 60 小时。

4. 多数受访者承担的教学任务较少，近半数受访者每周教学工作时长在 0-4 小时之间。

5. 受访者认为，阻碍其做出好的科研成果的最主要原因是——缺乏高质量、成体系的科研基础设施。

6. 学术界受访者的研究方向主要包括机器学习、计算机视觉、自然语言处理、信息检索与挖掘、推荐算法等。

7. 学术界受访者中最常用的深度学习算法是卷积神经网络（CNN），其次是循环神经网络（RNN、LSTM、GRU）和 Transformer。

8.在预训练研究领域，学术界最感兴趣的是研究知识和常识与模型的融合，其次是多模态数据的学习。

9.超过六成的学术界受访者最需要获得的支持是专业交流和人脉圈。其次是行业信息资讯。

(三) 工业界调研结论

1.工业界受访者的职业主要包括算法工程师、软件工程师、硬件工程师等，也有少部分受访者从事商业分析、产品经理等工作。

2.多数工业界受访者的 AI 研发年限在 5 年以下，具有 10 年以上 AI 研发经验的受访者非常少。

3.工业界受访者的月薪主要集中在 10000-19999 元和 20000-49999 元两个区间。

4.多数工业界受访者的周工作时长少于 50 小时，但有超过五分之一受访者正处于“996”或更长时间的加班状态中。

5.超过三分之一工业界受访者从事 IT、互联网等行业。从事公共管理、社会保障和社会组织、科学研究和技术服务、制造业等行业的受访者数量也比较多。

6.七成以上的工业界受访者所在的单位是私营企业，在外企工作的受访者数量次之。

7. 工业界研发团队的规模集中在 50-249 人和 0-49 人两个区间。但也有个别受访者所在的研发团队规模超过万人。

8.超过 99%的受访者所在企业使用 AI, 其中最常见的方式是使用机器学习实现预测分析和决策优化, 其次是在业务中使用 AI 系统实现创新和提升效率(降本增效)。

9.在工业界, 除了计算机视觉和自然语言处理外, 有较多的受访者从事推荐算法、信息检索与挖掘等研究方向。从事机器学习的受访者数量仅排第五。

10.除 GitHub 等渠道外, 技术博客也是工业界受访者获取 AI 信息的重要途径。

11.近半数工业界受访者希望从 AI 社区获得学术活动, 其次是知识课程。论文解读的占比仅排第三。

(四) 投资界调研结论

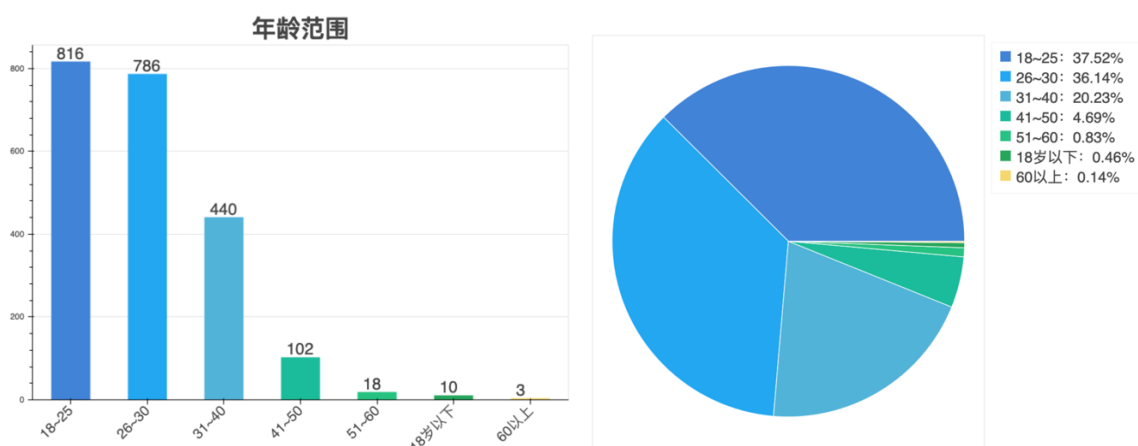
1.创业选择上, 投资界的受访者最看好 AI 芯片和智能制造两个赛道。

2.投资选择上, 投资界的受访者最看好智能制造和智能金融赛道, 而选择 AI 芯片的受访者数量仅排第三。

一、受访者总体情况

本次调查时间持续大约三周，从 2021 年 11 月 22 日起，截至 2021 年 12 月 15 日。在此期间，共收到问卷回复 9036 份。

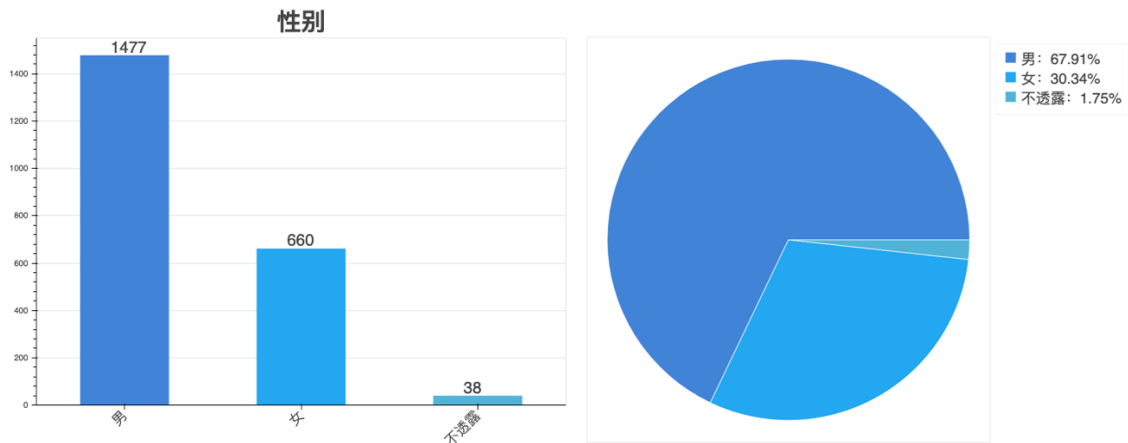
1. 年龄分布



统计中，受访者主要集中在 30 岁以下，其中 18-26 岁的数量最多，占比达到 37.52%。其次是 26-30 岁的受访者占比 36.14%，二者之和占比超过了 70%。31-40 岁的占比超过 20%，41-50 岁的占比不到 5%。

2. 性别分布

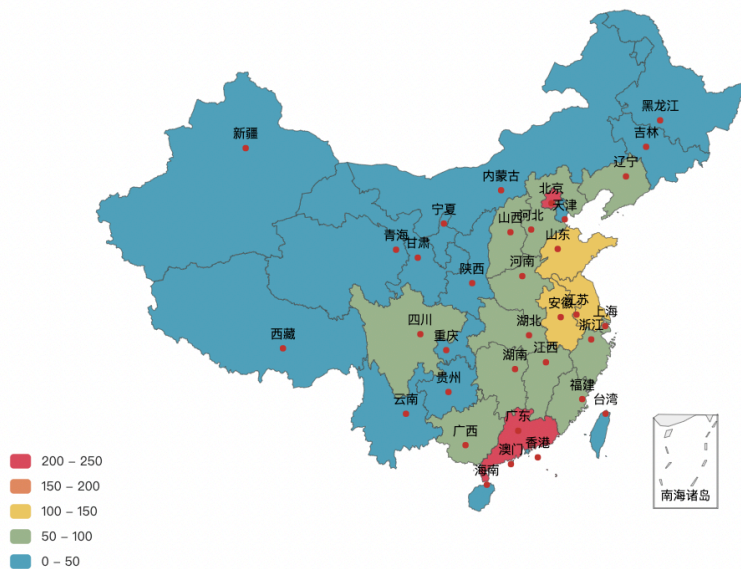
性别方面，受访者中的男性数量明显高于女性。男性受访者共计 1477 人，占比接近 70%。女性受访者共计 660 人，占比约 30%。



3.地区分布

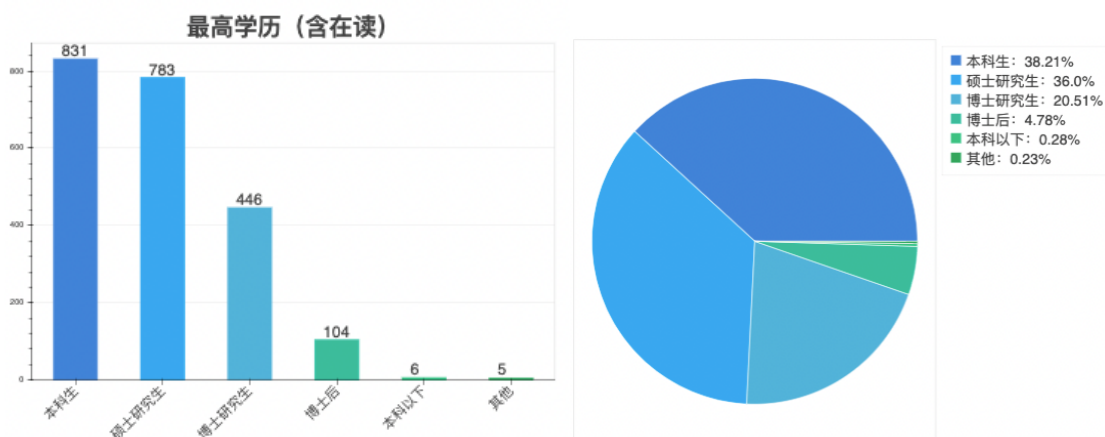
按受访者填写问卷的 IP 地址进行统计，在国内的受访者数量远大于海外。海外受访者占比不到一成，他们来自美国、英国、日本、韩国、新加坡等 12 个国家和地区。国内回复问卷数量较多的省级行政单位包括北京、广东、山东、江苏、浙江、上海等。其中来自北京的受访者数量最多，占比接近二成。

国内回复情况



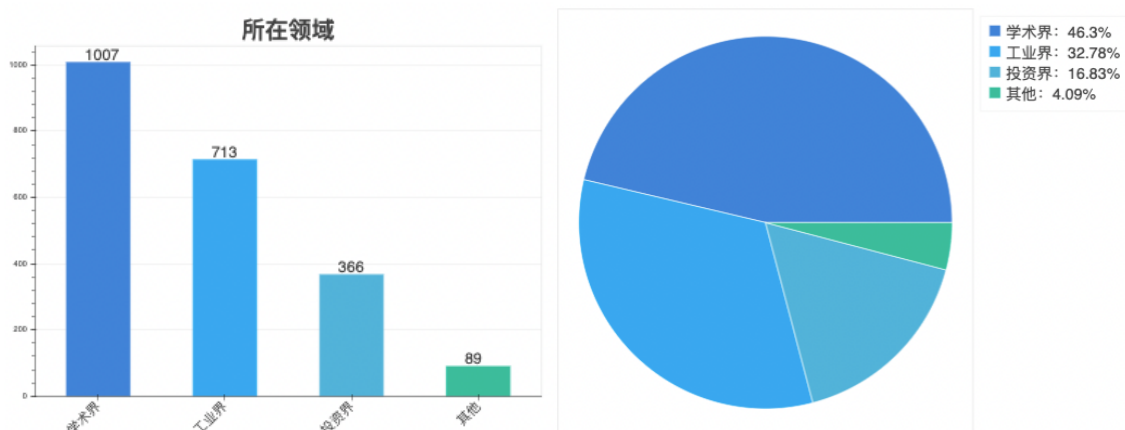
4.最高学历分布

受访者的普遍拥有本科及以上学历学位，占比达到九成以上。其中，最高学历（含在读）为本科受访者数量最多，有 831 人，占比接近半数，接着是硕士研究生和博士研究生，总占比超过半数。



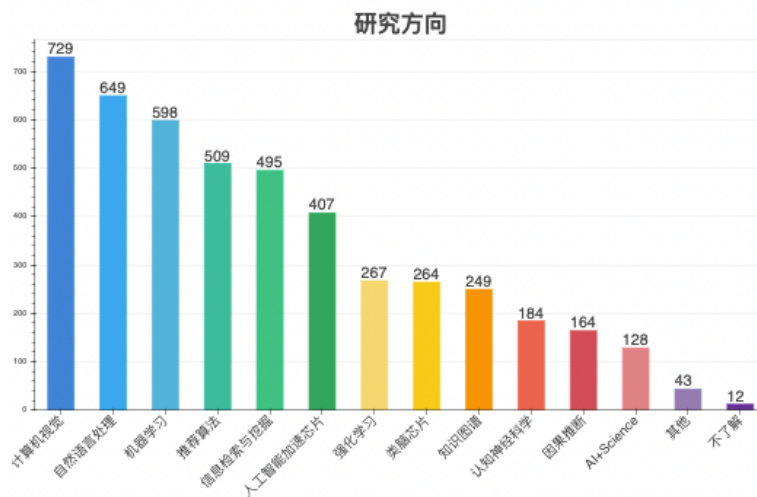
5.所在领域分布

受访者主要来自学术界和工业界，其中学术界受访者占比近半，工业界受访者占比超过三成。投资界受访者仅为学术界受访者的三分之一。



6.AI 研究方向

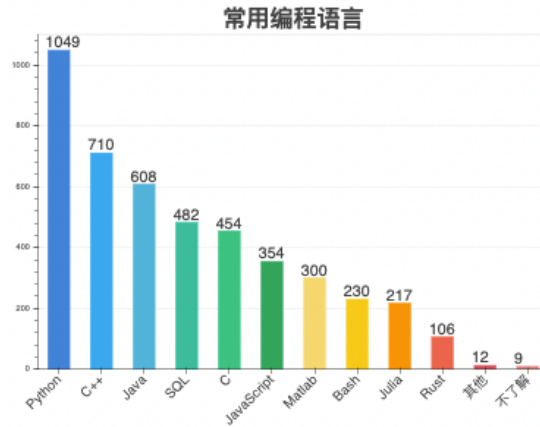
计算机视觉是从事研究人数最多的方向，在受访者中的占比超过 30%，其次是自然语言处理，在受访者中占比与计算机视觉相近。机器学习和推荐算法的研究人数分别排名第三、第四，第五是信息检索与挖掘方向。



选项 (Top5)	占比 (选项为多选)
计算机视觉	33.52%
自然语言处理	29.84%
机器学习	27.49%
推荐算法	23.4%
信息检索与挖掘	22.76%

7.编程语言

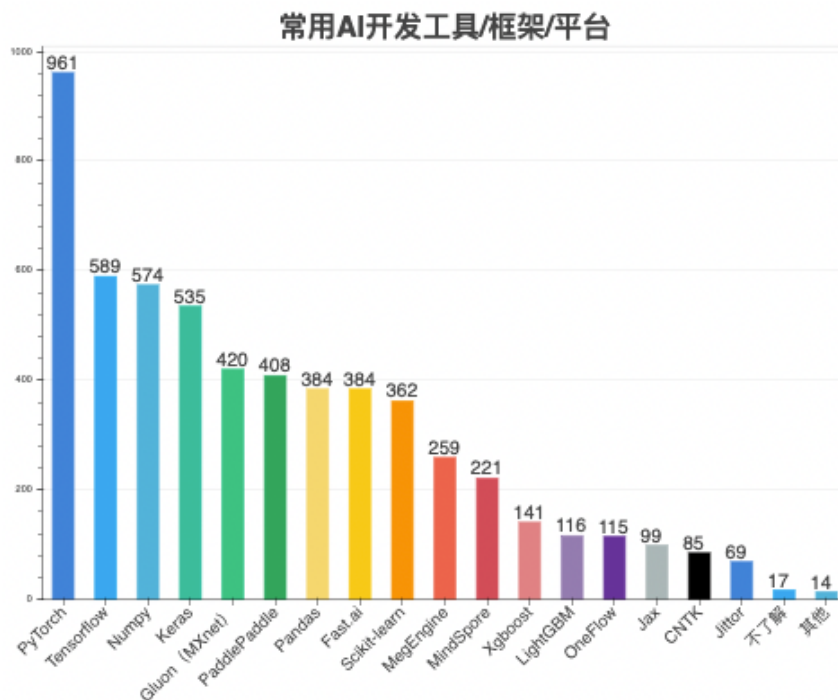
Python 是受访者中最常用的编程语言，占比接近 50%，其次是 C++和 Java。



选项 (Top5)	占比 (选项为多选)
Python	48.23%
C++	32.64%
Java	27.95%
SQL	22.16%
C	20.87%

8.常用的 AI 开发工具/框架/平台

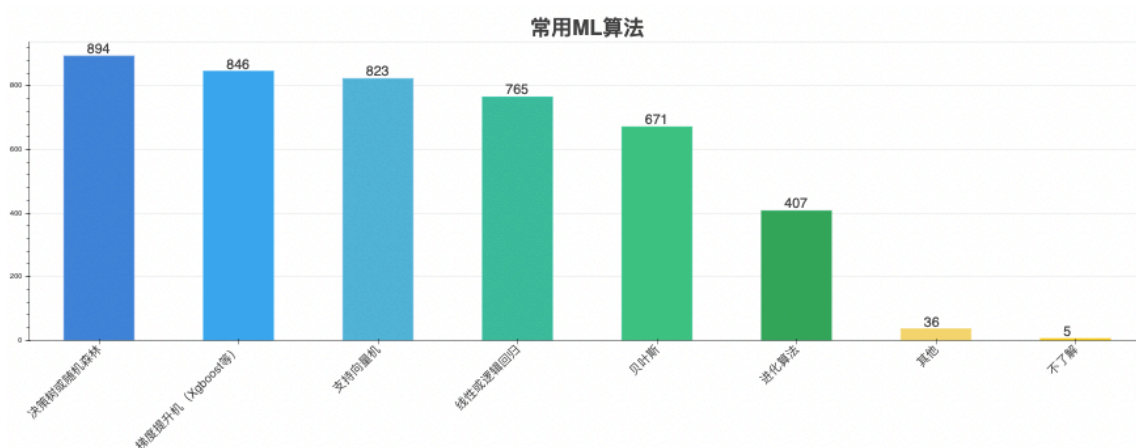
PyTorch 是受访者中使用最广泛的 AI 开发框架，占比超过了 40%，几乎是 TensorFlow 用户数量的 1.5 倍。随后是 TensorFlow 和 Numpy，占比均在 25% 以上。



选项 (Top5)	占比 (选项为多选)
PyTorch	44.18%
TensorFlow	27.08%
Numpy	26.39%
Keras	24.60%
Gluon (MXnet)	19.31%

9.常用的机器学习算法

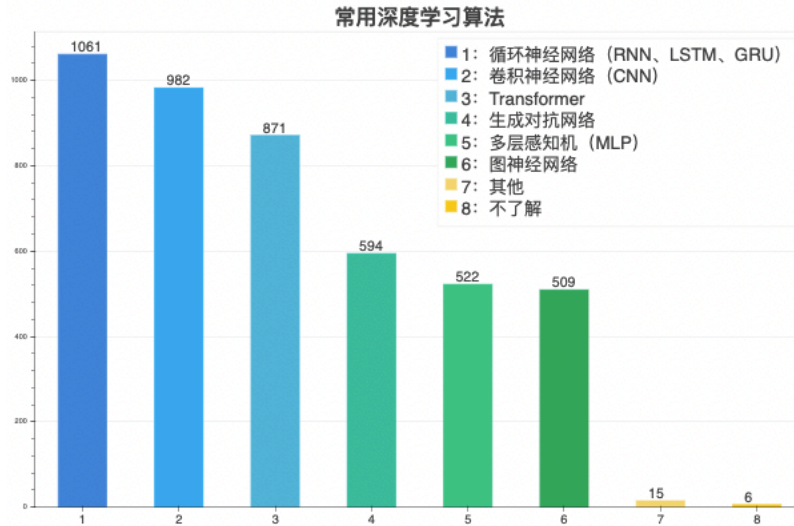
在常用的机器算法方面，受访者最常使用的是决策树或随机森林算法（41.1%），其次是梯度提升机（38.9%）。支持向量机和线性或逻辑回归算法的使用者数量也较多，在受访者中二者的占比分别为 37.84%和 35.17%。



选项 (Top5)	占比 (选项为多选)
决策树或随机森林	41.1%
梯度提升机 (Xgboost 等)	38.9%
支持向量机	37.84%
线性或逻辑回归	35.17%
贝叶斯	30.85%

10.常用的深度学习算法

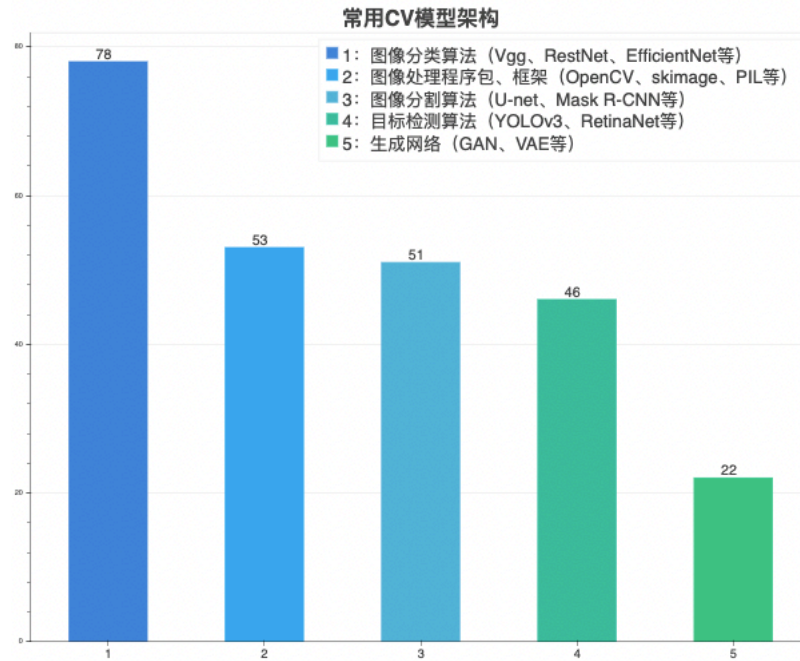
循环神经网络（如 RNN，LSTM 和 GRU 等）是受访者中最常用的深度学习算法，将近 50%的受访者经常使用。其次是 CNN，在受访者中的占比超过了 40%。使用 Transformer 的受访者数量排第三，占比也超过 40%。



选项 (Top5)	占比 (选项为多选)
循环神经网络 (RNN、LSTM、GRU 等)	48.78%
卷积神经网络 (CNN)	45.15%
Transformer	40.05%
生成对抗网络	27.31%
多层感知机 (MLP)	24.0%

11. 计算机视觉领域常用架构

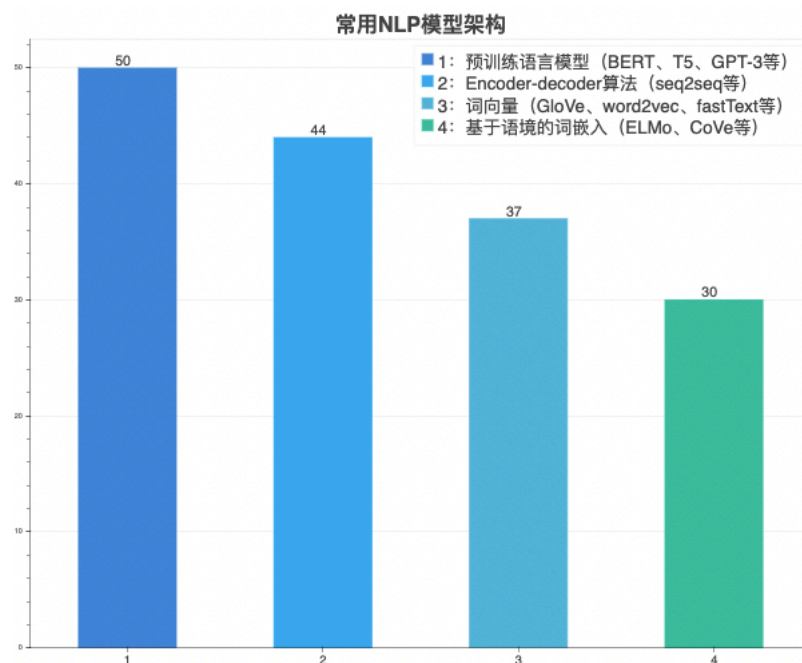
在选择研究方向为“计算机视觉”的受访者中，图像分类算法是目前最常用的架构，58.65%的受访者经常使用。



选项 (Top5)	占比 (选项为多选)
图像分类算法 (Vgg、RestNet、EfficientNet 等)	58.65%
图像处理程序包、框架 (OpenCV、skimage、PIL 等)	39.85%
图像分割算法 (U-net、Mask R-CNN 等)	38.35%
目标检测算法 (YOLOv3、RetinaNet 等)	34.59%
生成网络 (GAN、VAE 等)	16.54%

12. 自然语言处理领域常用架构

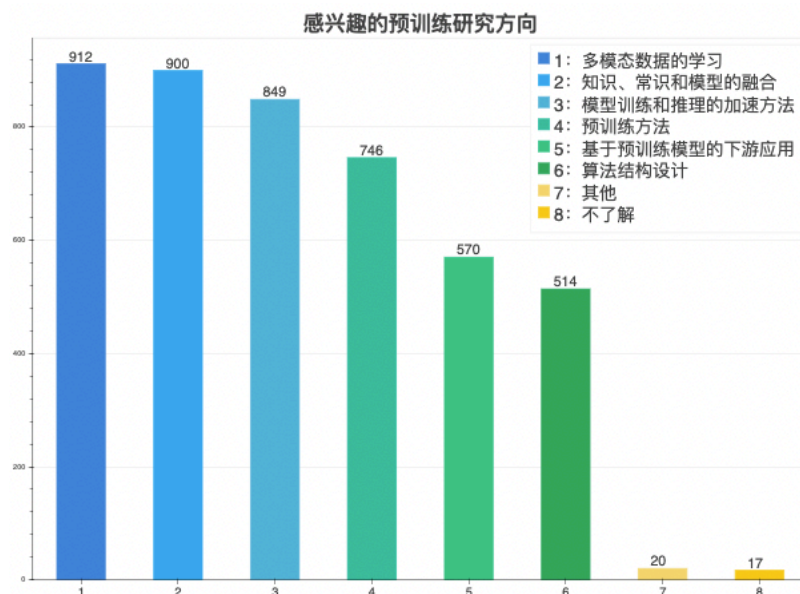
在自然语言处理方向的受访者中，预训练模型是当前使用人数最多的模型架构，占比接近60%。



选项 (Top4)	占比 (选项为多选)
预训练语言模型 (BERT、T5、GPT-3 等)	58.14%
Encoder-decoder 算法 (seq2seq 等)	51.16%
词向量 (GloVe、word2vec、fastText 等)	43.02%
基于语境的词嵌入 (ELMo、CoVe 等)	34.88%

13.感兴趣的预训练研究方向

受访者对于预训练领域的多个方面展现出研究的兴趣。目前最受关注的是多模态数据的学习，占比超过 40%，其次是知识、常识和模型的融合。



选项 (Top5)	占比 (选项为多选)
多模态数据的学习	41.93%
知识、常识和模型的融合	41.38%
模型训练和推理的加速方法	39.03%
预训练方法	34.3%
基于预训练模型的下游应用	26.21%

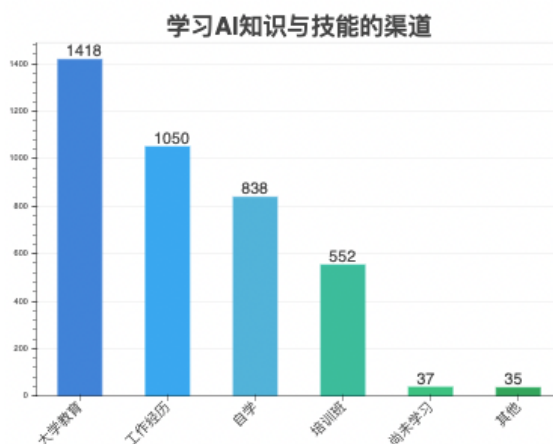
14.2021 年十大 AI 热点

受访者票选的 2021 年十大 AI 热点如下所示。其中 Angrist、Imbens 在因果关系分析方法论中的研究获得诺贝尔奖夺得第一。其次是 OpenAI 发布多模态预训练模型 DALL·E 和 CLIP，谷歌的 Switch Transformer 排名第三。从排名中可以看到，和预训练模型（大模型、基础模型等）相关的事件较多，说明预训练在 2021 年是最受关注的研究领域之一。

序号	选项 (Top10)	票数 (多选)
1	David Card, Joshua D. Angrist 和 Guido W. Imbens 因对因果关系分析方法论的贡献获得诺贝尔经济学奖	795
2	OpenAI 发布多模态预训练模型 DALL·E 和 CLIP	749
3	谷歌发布 1.6 万亿参数模型 Switch Transformer	587
4	智源研究院发布万亿规模超大规模智能模型悟道 2.0	457
5	AlphaFold2/RosTTAFold 算法和相关数据集等开源	428
6	微软亚洲研究院提出 Swin Transformer 视觉预训练模型	333
7	何恺明团队提出 Masked Autoencoders 视觉预训练方法	330
8	清华、智源等提出 Pre-Train Prompt Tuning (PPT) 框架	302

9	斯坦福大学研究者利用 AI 和脑机接口实现“意念写字”	286
10	谷歌发布改进版视觉预训练模型 ViT	259

15.学习 AI 知识与技能的渠道



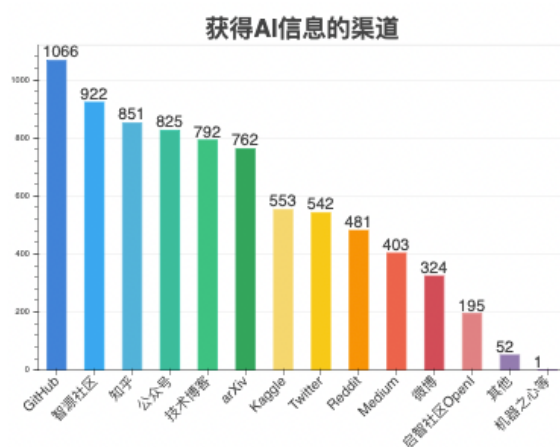
选项	占比 (选项为多选)
大学教育	65.2%
工作经历	48.28%
自学	38.53%
培训班	25.38%
尚未学习	1.7%
其他	1.61%

受访者学习 AI 知识与技能的主要两个渠道是大学教育和工作经历，分别有 65.2%和 48.28%的受访者通过大学教育和工作经历获得相关的知识和技能。但是，对于受访者而言，通过自学和培训班学习，是重要的补充学习渠道，将近 40%的受访者采用自学的方式获取相关知识和技能，采用培训班渠道的受访者也超过了 25%。没有学习过 AI 知识的受访者的数量很少。

16. 获得 AI 信息的渠道

受访者中，获得 AI 信息主要渠道包括 GitHub、智源社区、知乎、公众号和技术博客等。

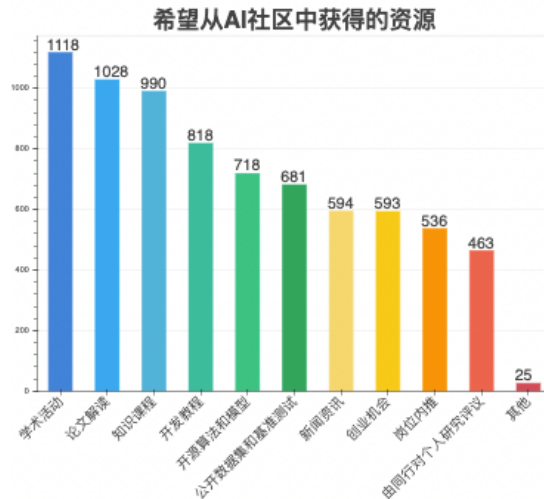
其中，将近 50% 的受访者从 GitHub 获得 AI 信息。而智源社区是受访者第二大获取 AI 信息的渠道，超过 40% 受访者使用这一平台。



选项 (Top5)	占比 (选项为多选)
GitHub	49.01%
智源社区	42.39%
知乎	39.13%
公众号	37.93%
技术博客	36.41%

17. 希望从 AI 社区中获得的资源

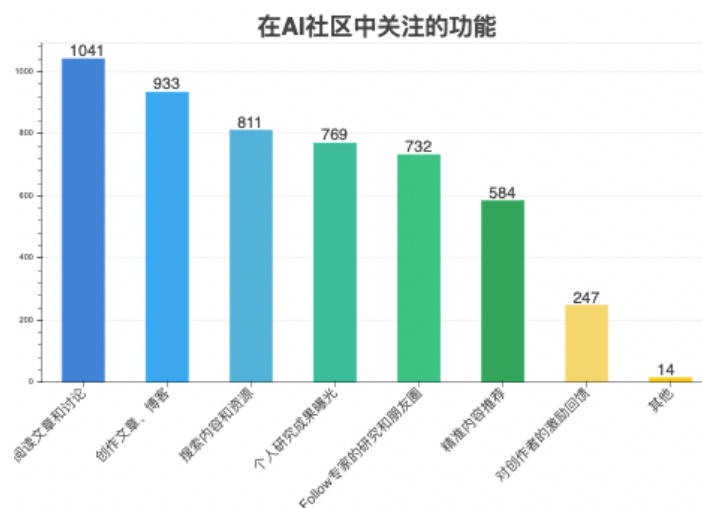
受访者希望从 AI 社区中获得的资源主要包括学术活动、知识课程、论文解读和开发教程等。其中，51.4% 的受访者希望获得学术活动资源，47.26% 希望获得论文解读资源，45.52% 希望获得知识课程。



选项 (Top5)	占比 (选项为多选)
学术活动	51.4%
论文解读	47.26%
知识课程	45.52%
开发教程	37.61%
开源算法与模型	33.01%

18.在 AI 社区中关注的功能

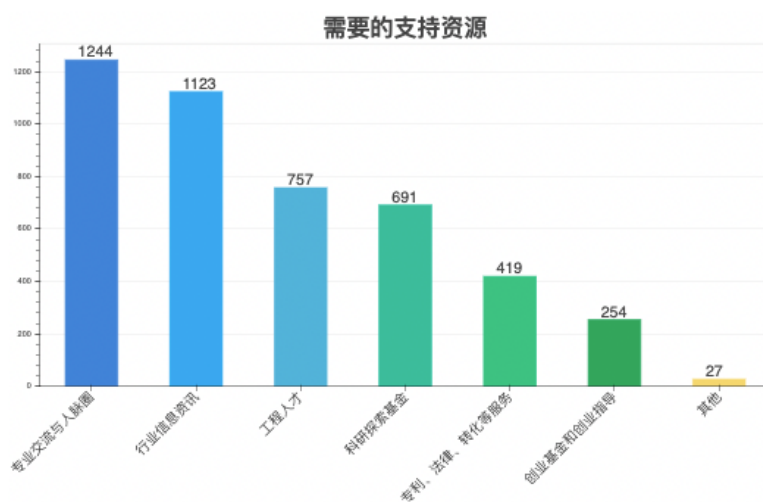
受访者在 AI 社区中最关注的两大核心功能是阅读 AI 相关的文章和讨论，以及创作文章、博客等内容。选择这两项功能的受访者占比都超过了 40%。



选项 (Top5)	占比 (选项为多选)
阅读文章和讨论	47.86%
创作文章、博客	42.90%
搜索内容和资源	37.29%
个人研究成果曝光	35.36%
Follow 专家的研究和朋友圈	33.66%

19.需要的支持资源

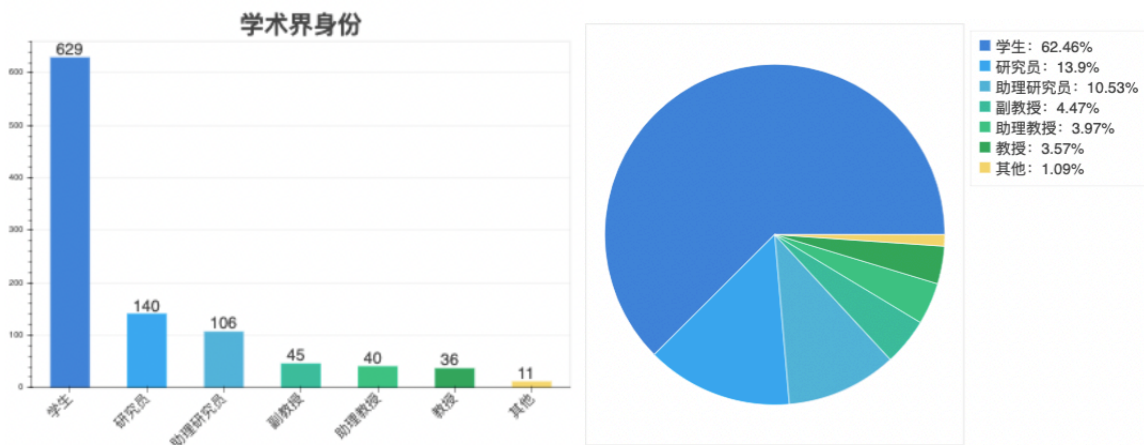
调查显示, 受访者最需要的支持是来自 AI 领域的专业交流和人脉圈, 选择此项的受访者占比接近 60%。其次是行业信息资讯, 占比超过 50%。



选项	占比 (选项为多选)
专业交流与人脉圈	57.20%
行业信息资讯	51.63%
工程人才	34.80%
科研探索基金	31.77%
专利、法律、转化等服务	19.26%
创业基金和创业指导	11.68%
其他	1.24%

二、学术界受访者情况

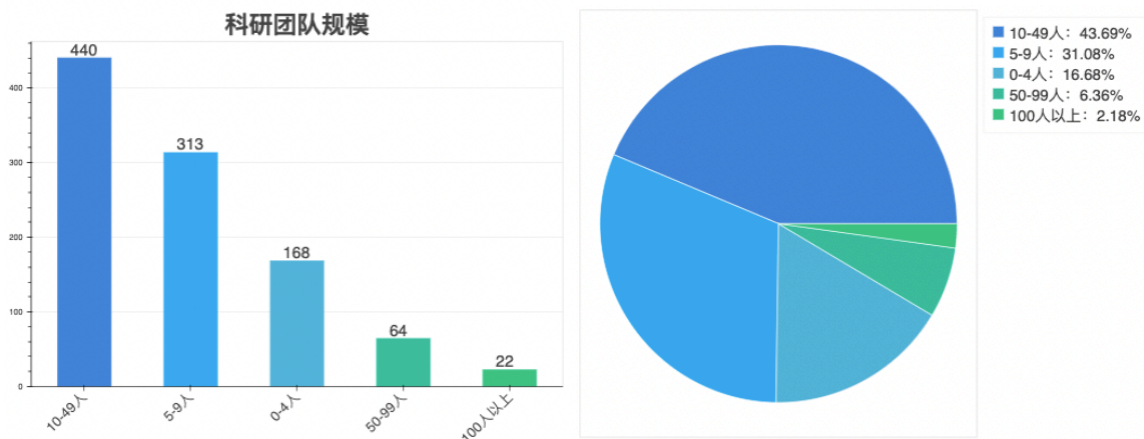
1. 学术界身份



学术界中，超过 60% 受访者身份是学生。数量仅次于学生的是研究员（13.9%）和助理研究员（10.53%）。学生数量是有教职研究者数量的 5 倍左右。

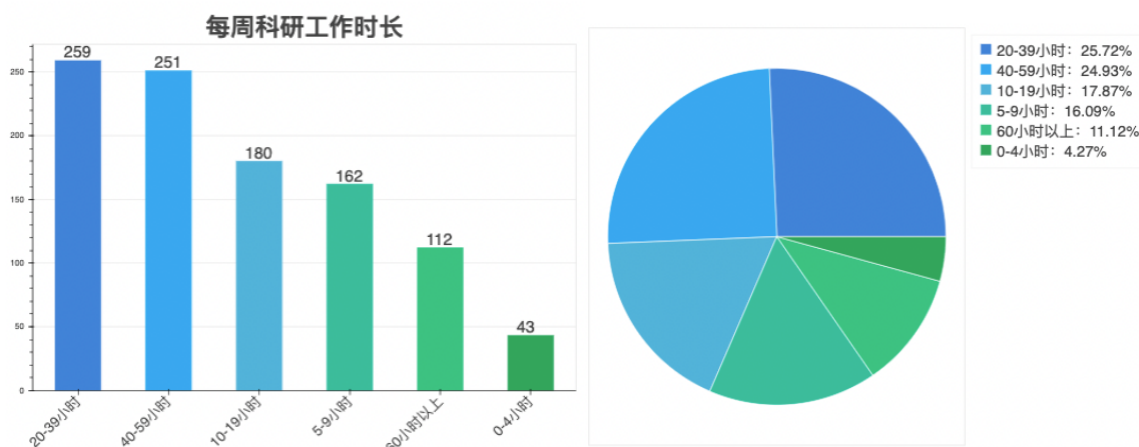
2. 科研团队规模

学术界中，43.69% 的受访者所在的 AI 团队规模在 10-49 人之间。超过 100 人的科研团队在受访者中较为少见，仅占 2.18%。



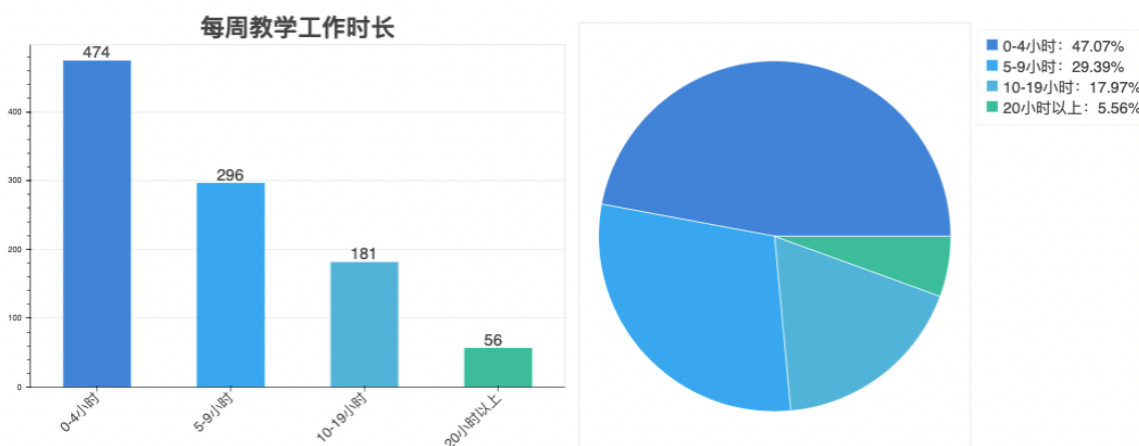
3.每周科研工作时长

科研时长方面，超过四分之一的受访者每周科研时间在 20-39 小时之间，相当于平均个工作日科研 5-8 个小时。其次是每周科研时长 40-59 小时的受访者，相当于平均每个工作日科研时长在 8-12 个小时。甚至有 11.12%的受访者每周科研工作超过 60 小时。



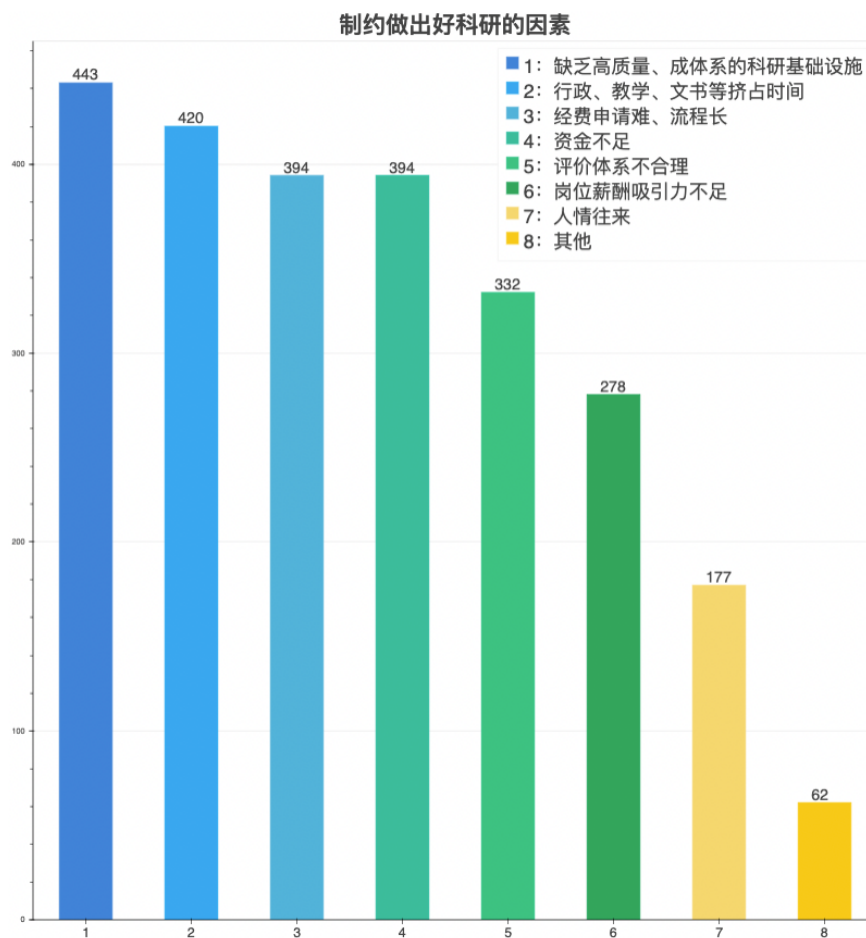
4.每周教学工作时长

教学时长方面，多数受访者承担的教学任务较少。将近 50%的受访者每周教学工作时长在 0-4 小时。只有不到 25%的受访者每周教学工作时长超过 10 小时。



5.阻碍做出好科研的因素

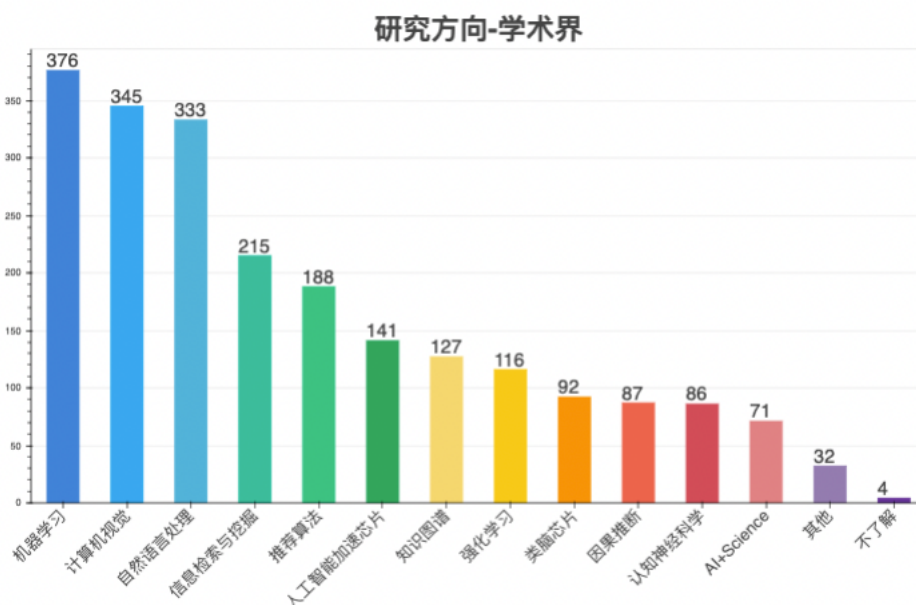
学术界的受访者认为，缺乏高质量、成体系的科研基础设施（如软硬件平台、工具等）是阻碍科研的最主要因素。有 43.99%的受访者选择这两项。此外，非科研事务挤占时间也被认为是阻碍做出好科研的重要因素，选择这项的受访者占比为 41.71%。



选项 (Top5)	占比 (选项为多选)
缺乏高质量、成体系的科研基础设施	43.99%
行政、教学、文书等挤占时间	41.71%
经费申请难、流程长	39.13%
资金不足	39.13%
评价体系不合理	32.97%

6.学术界受访者的研究方向

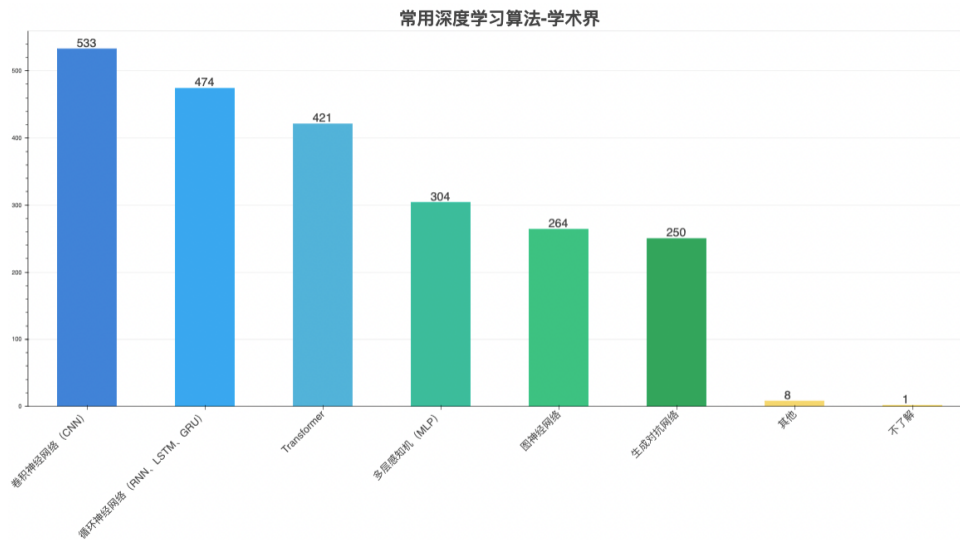
学术界受访者中，从事机器学习的受访者数量最多，占比为 37.34%，接着是从事计算机视觉和自然语言处理的受访者，占比分别为 34.26%和 33.07%。



选项 (Top5)	占比 (选项为多选)
机器学习	37.34%
计算机视觉	34.26%
自然语言处理	33.07%
信息检索与挖掘	21.35%
推荐算法	18.67%

7.学术界常用的深度学习算法

与总体趋势不同，在学术界的受访者中，最常使用的深度算法是卷积神经网络（CNN），在受访者中的占比超过 50%。其次是循环神经网络(RNN、LSTM、GRU)，占比为 47.07%。

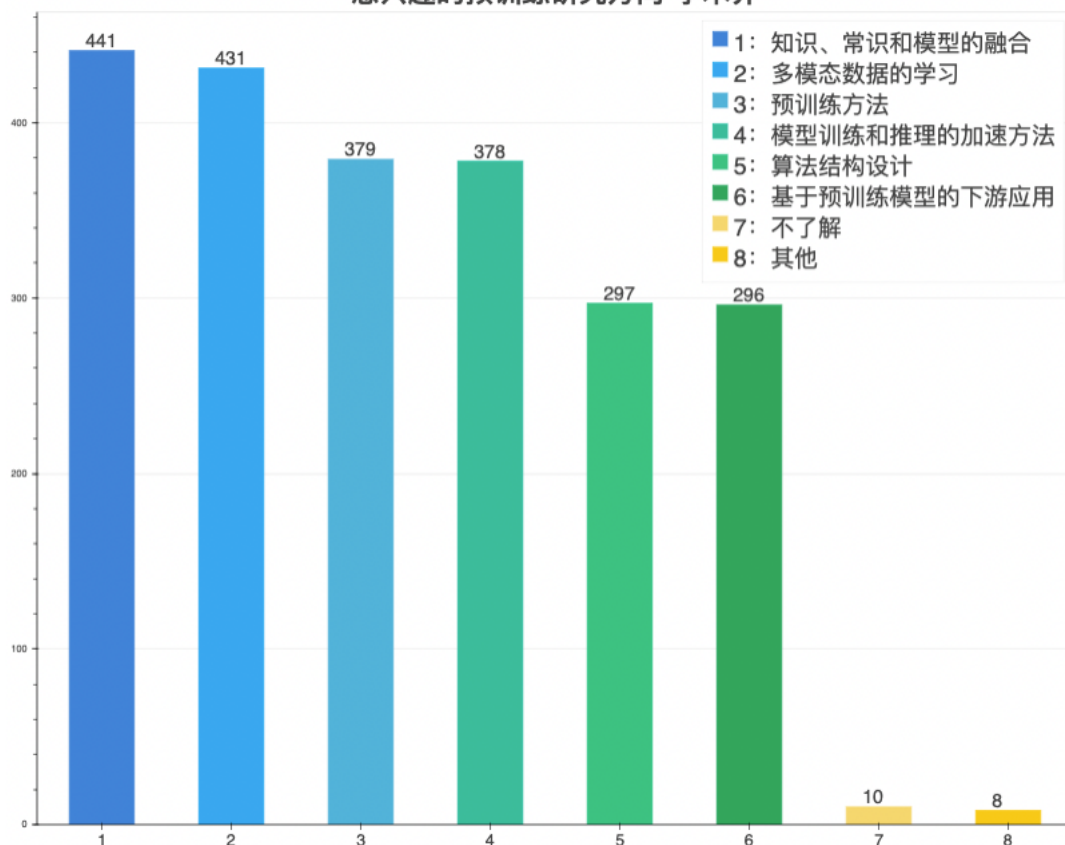


选项 (Top5)	占比 (选项为多选)
卷积神经网络 (CNN)	52.93%
循环神经网络 (RNN、LSTM、GRU)	47.07%
Transformer	41.81%
多层感知机 (MLP)	30.19%
图神经网络	26.22%

8.学术界感兴趣的预训练方向

与总体趋势不同，在预训练领域，学术界受访者最关注的是将知识、常识等与模型融合，有 43.79%的受访者对此感兴趣。其次是多模态数据学习，占比为 42.8%。

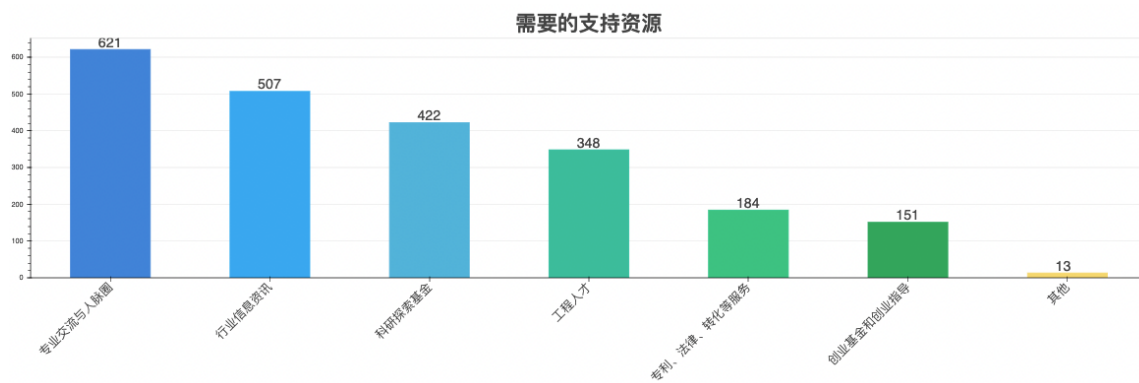
感兴趣的预训练研究方向-学术界



选项 (Top5)	占比 (选项为多选)
知识、常识和模型的融合	43.79%
多模态数据的学习	42.8%
预训练方法	37.64%
模型训练和推理加速方法	37.54%
算法结构设计	29.49%

9.学术界需要获得的支持

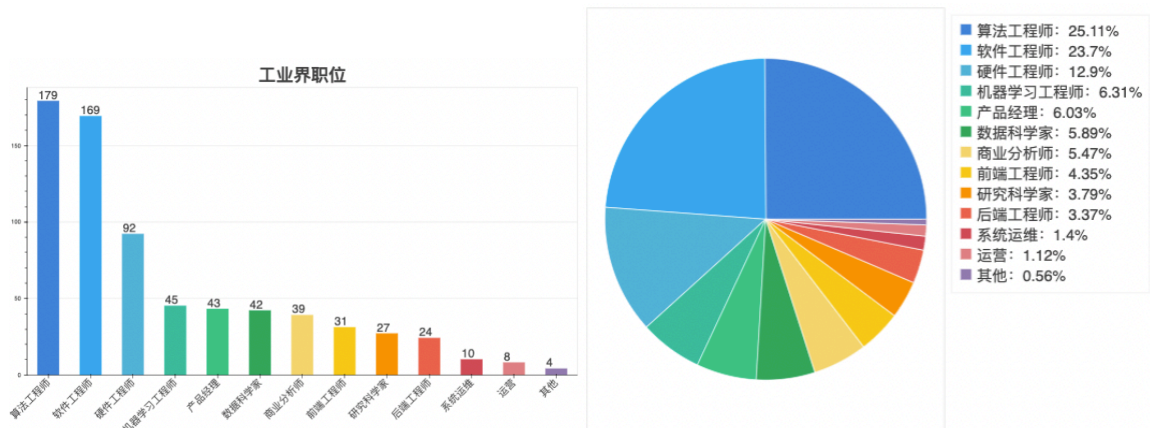
学术界的受访者中，超过六成期待获得专业交流和人脉圈等资源支持。其次是行业资讯和科研探索基金。



选项 (Top5)	占比 (选项为多选)
专业交流与人脉圈	61.67%
行业信息资讯	50.35%
科研探索基金	41.91%
工程人才	34.56%
专利、法律、转化等服务	18.27%

三、工业界受访者情况

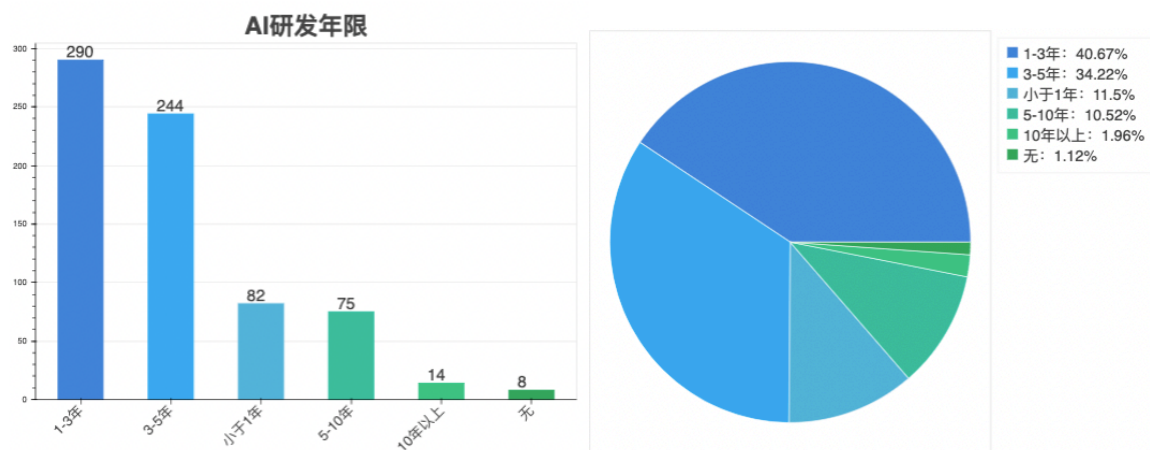
1. 工业界职位



工业界受访者中,数量最多的是算法工程师,占比达到 25.11%,其次是软件工程师(23.7%)和硬件工程师(12.9%),机器学习工程师在受访者中的占比为 6.31%。

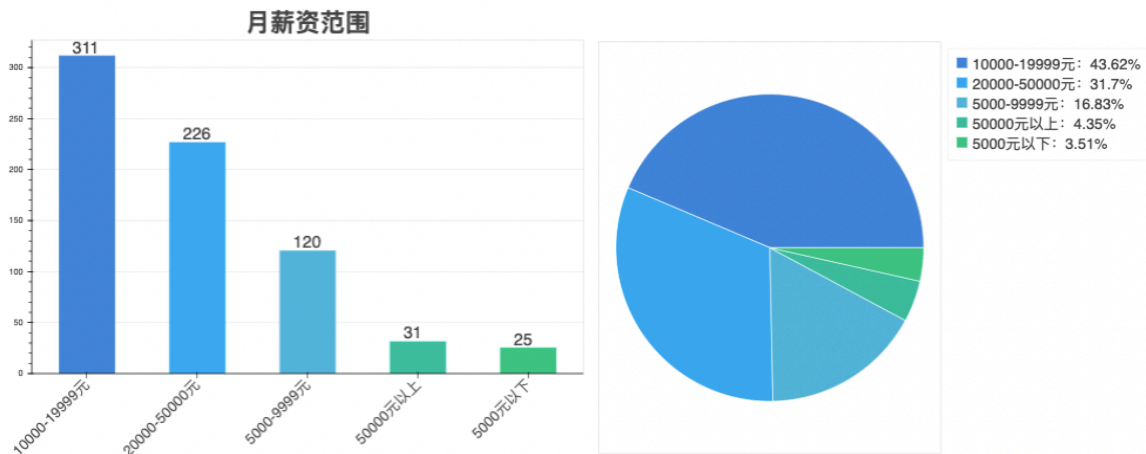
2. AI 研发年限

工业界中受访者的 AI 研发年限普遍在 1-3 年,这部分受访者占总数的 40.67%,其次是有 3-5 年经验的受访者,占比超过 30%。10 年以上 AI 研发经验的受访者仅占总数的 1.96%。



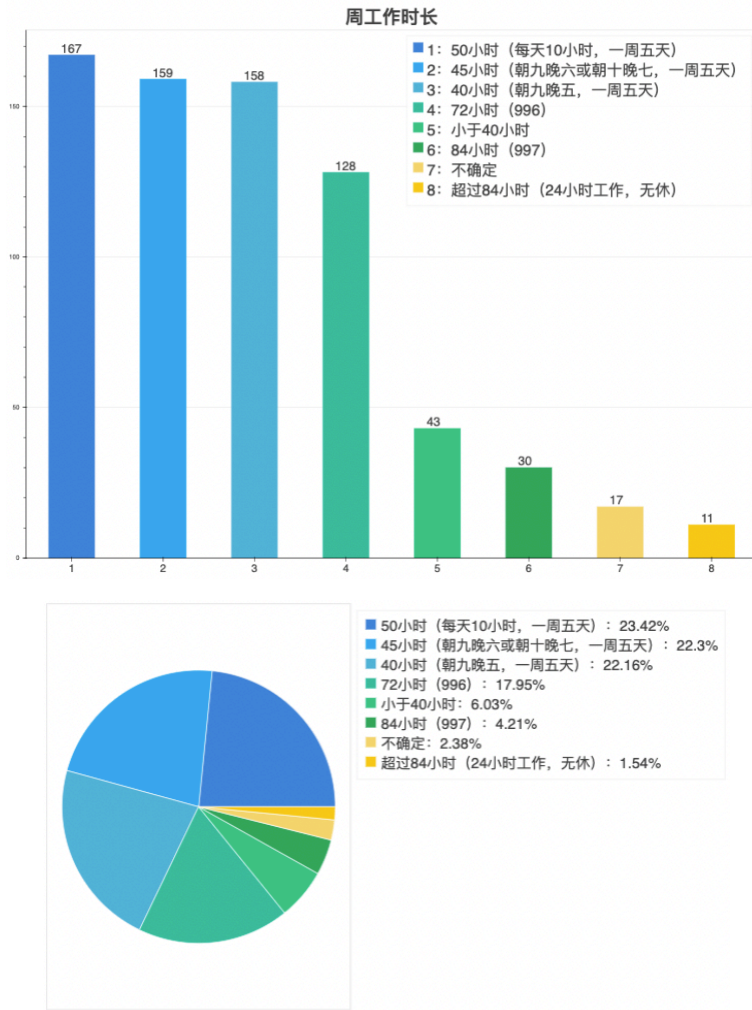
3.月薪资范围

工业界中，有超过 40%的受访者月薪在 10000-19999 元之间。同时，有超过 30%的受访者月薪在 20000-50000 元之间。5000 元以下和 50000 元以上薪资的受访者占比都较小。



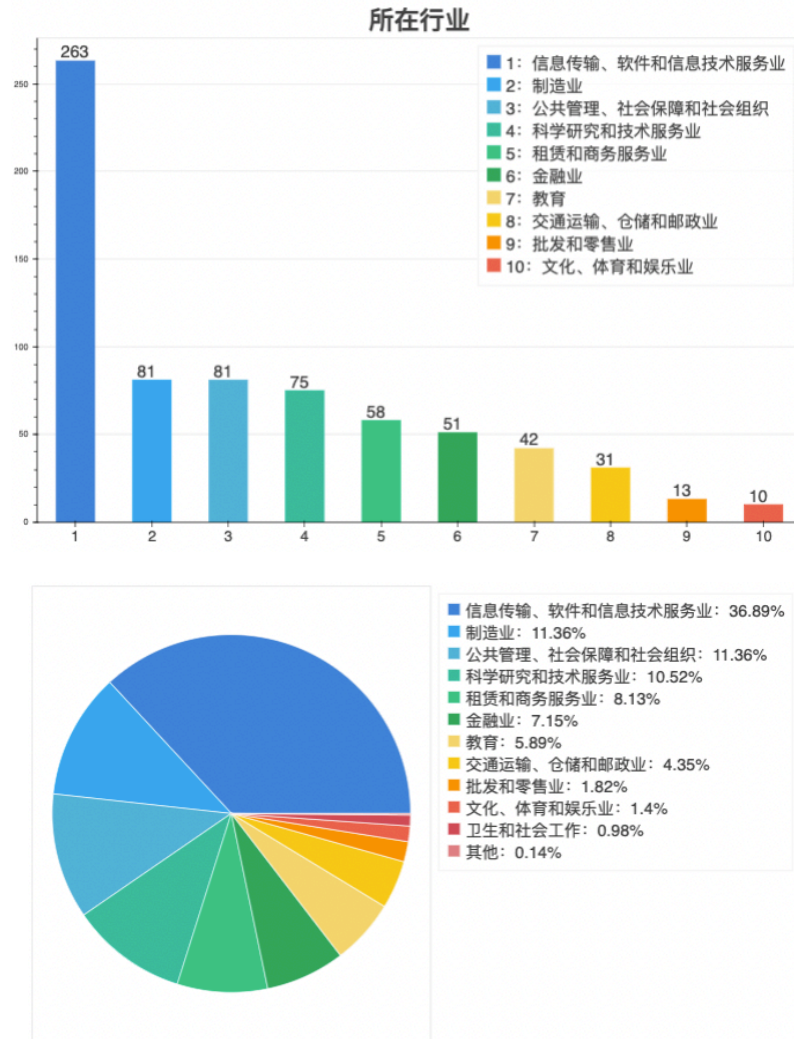
4.周工作时长

工业界受访者中，将近 25%的受访者每周工作 50 小时，接着是每周工作 45 小时和每周工作 40 小时的受访者，占比均在 22%左右。此外，有超过五分之一的受访者每周工作时长在 72 小时及以上。



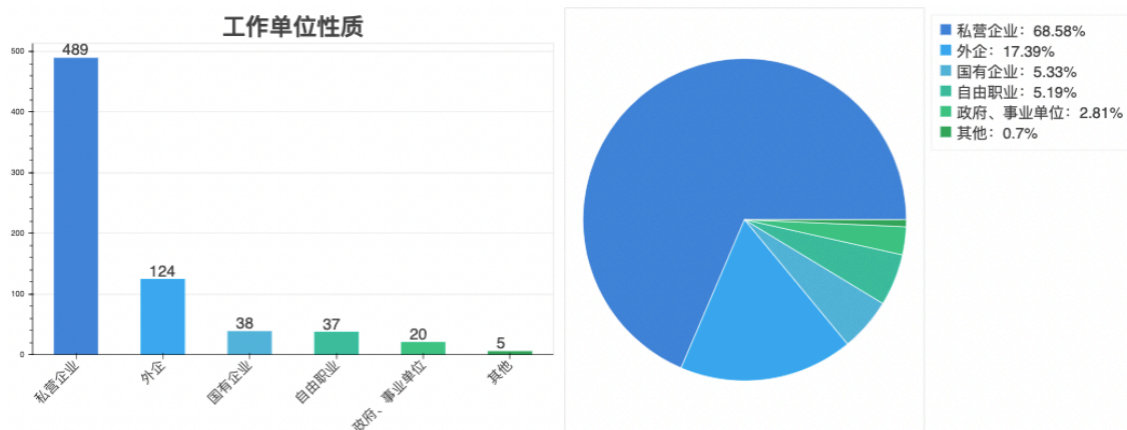
5.所在行业

行业方面，超过 30%的工业界受访者从事的是信息技术服务业，如互联网、IT 等。同时，从事制造业、公共管理、社会保障和社会组织，科学研究和技术服务等行业的受访者数量也相对较多，占比均超过 10%。



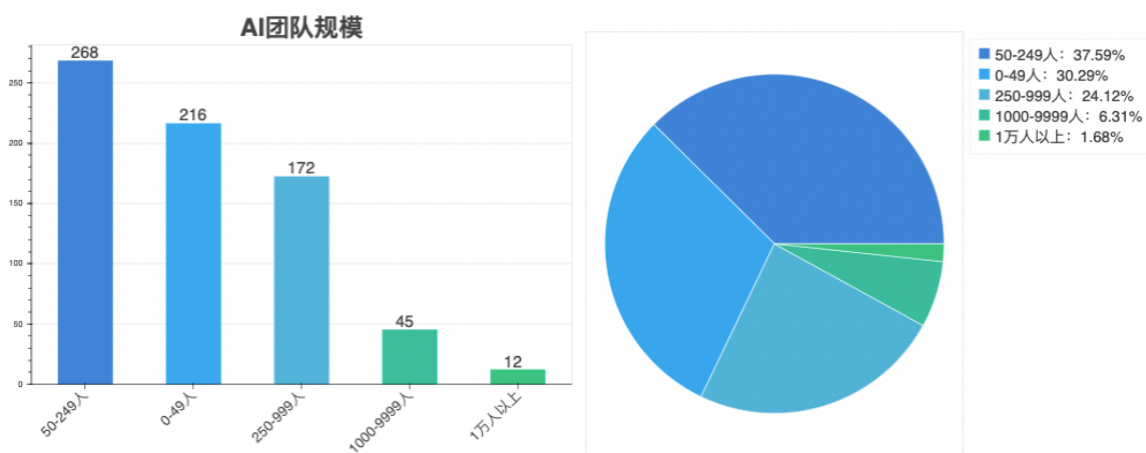
6. 工作单位性质

受访者中，将近 70% 以上的工业界受访者在私营企业工作。在外企工作的受访者数量次之，占比为 17.39%。



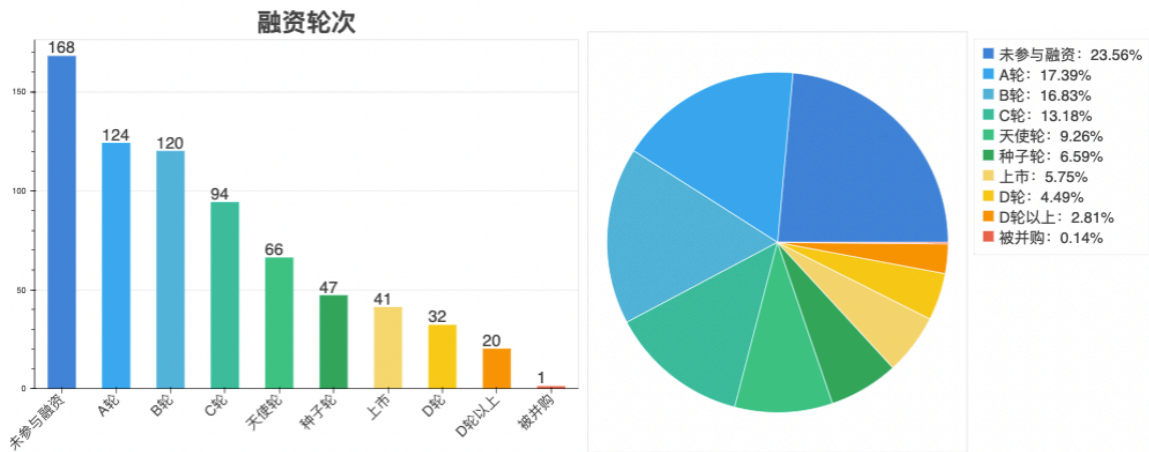
7.研发团队规模

接近 40% 工业界受访者所在的 AI 团队规模在 50-249 人之间，超过 30% 的受访者所在的 AI 团队规模在 0-49 人之间。不到 10% 的受访者所在的 AI 团队规模超过千人，所在 AI 团队超过万人的受访者仅有 1.68%。



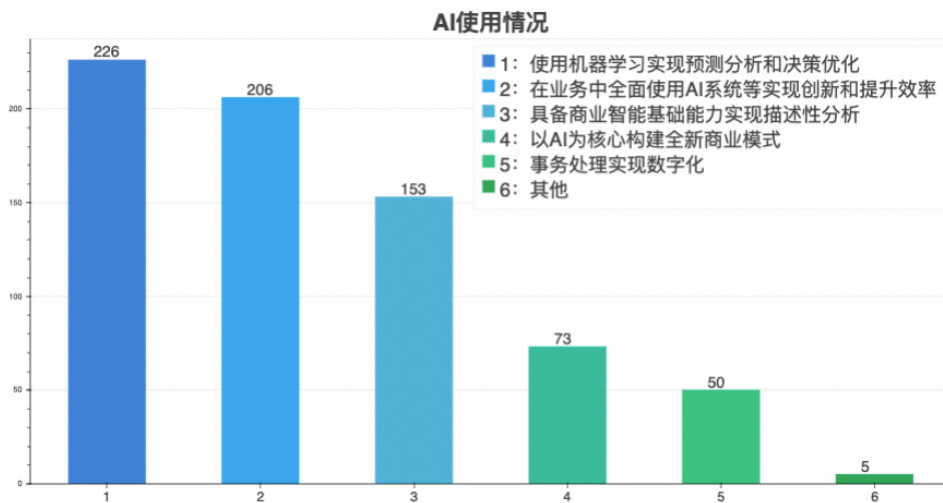
8.企业融资轮次

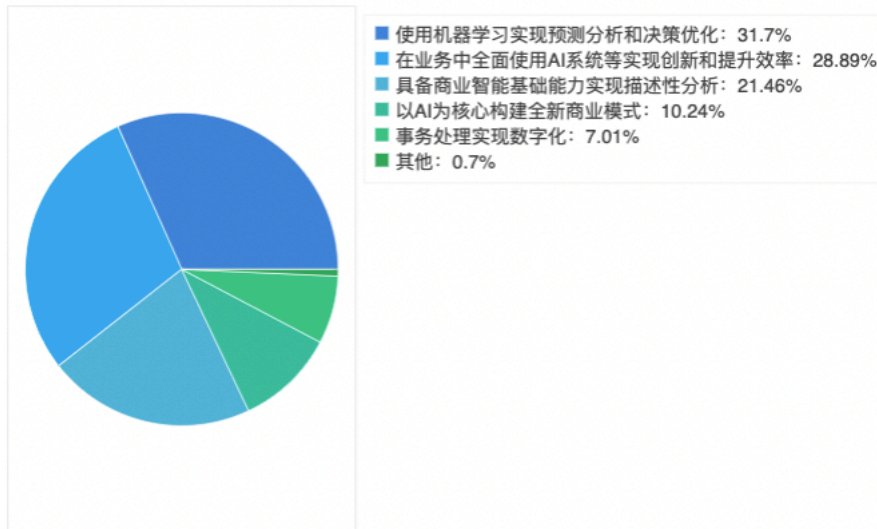
有超过 20% 的受访者所在企业未融资，超过 15% 的受访者所在企业经历了 A 轮或 B 轮融资。此外，受访者所在企业处于 D 轮及以上融资的不足 5%。



9.AI 在企业中的使用情况

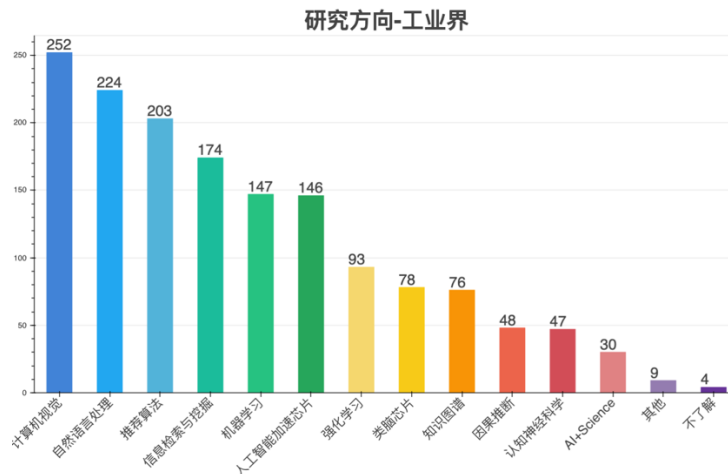
超过 99%的受访者所在的企业使用过 AI，其中最常见的方式是使用机器学习实现预测分析和决策优化（31.70%），其次是在业务中使用 AI 系统实现创新和提升效率（28.89%）。





10. 工业界受访者的研究方向

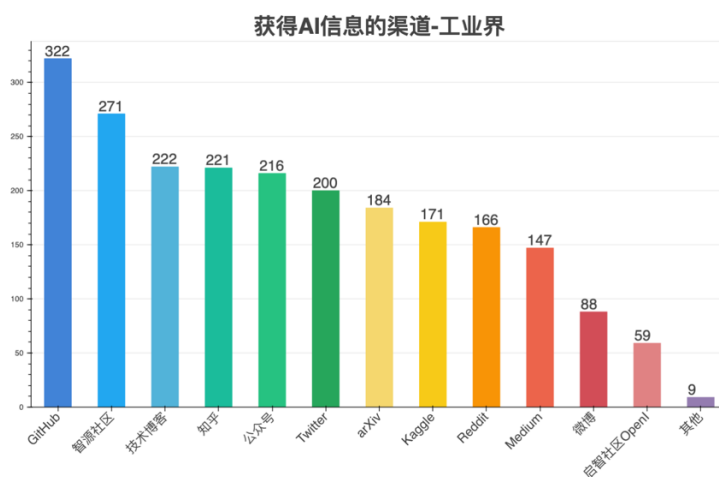
计算机视觉和自然语言处理是工业界中受访者主要关注的两大研究方向，从事这两个方向研究的占比分别为 35.34%和 31.42%。与总体趋势不同的是，工业界中研究机器学习受访者数量仅排第五。



选项 (Top5)	占比 (选项为多选)
计算机视觉	35.34%
自然语言处理	31.42%
推荐算法	28.47%
信息检索与挖掘	24.4%
机器学习	20.62%

11. 获得 AI 信息的渠道

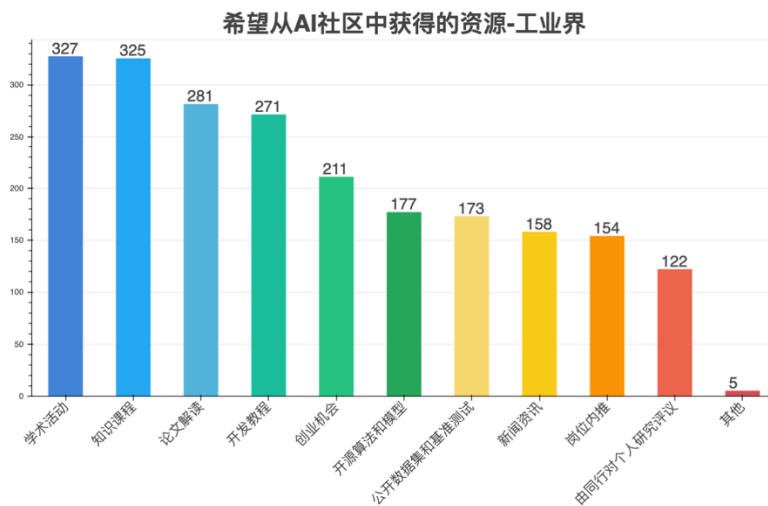
除了 GitHub 和智源社区以外，技术博客是工业界受访者获得 AI 信息的重要渠道之一，占比接近三分之一。



选项 (Top5)	占比 (选项为多选)
GitHub	45.16%
智源社区	38.01%
技术博客	31.14%
知乎	31.00%
公众号	30.29%

12. 希望从 AI 社区中获得的资源

与总体趋势不同，45.86%的工业界受访者希望从 AI 社区获得学术活动。其后是知识课程（45.58%）。论文解读的占比仅排第三。

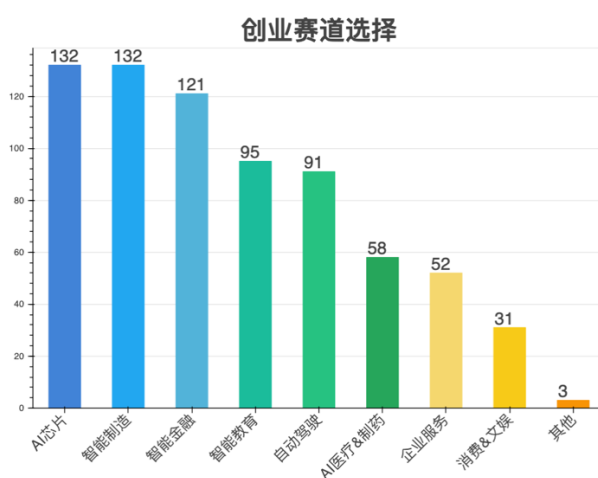


选项 (Top5)	占比 (选项为多选)
学术活动	45.86%
知识课程	45.58%
论文解读	39.41%
开发教程	38.01%
创业机会	29.59%

四、投资界受访者情况

1. 创业选择

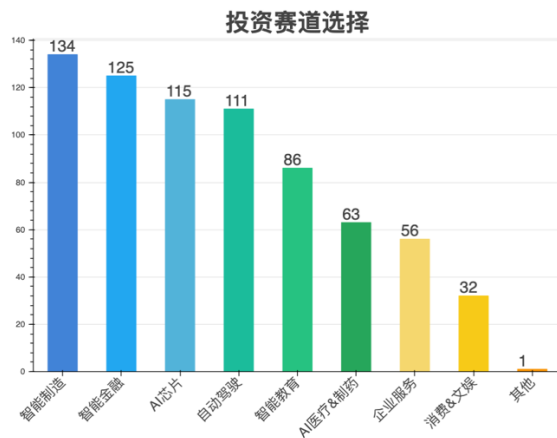
创业方面，投资界的受访者最看好 AI 芯片和智能制造两个赛道，分别有 36.07% 的投资界受访者选择了这两个赛道。



选项 (Top5)	占比 (选项为多选)
AI 芯片	36.07%
智能制造	36.07%
智能金融	33.06%
智能教育	25.96%
自动驾驶	24.86%

2. 投资选择

投资方面，投资界的受访者最看好智能制造和智能金融赛道，选择二者的受访者占比分别是 36.61% 和 34.15%。而选择 AI 芯片的受访者占比仅排第三。



选项 (Top5)	占比 (选项为多选)
智能制造	36.61%
智能金融	34.15%
AI 芯片	31.42%
自动驾驶	30.33%
智能教育	23.5%

五、研究方法

(一) 问卷题目

本调研采用网络问卷形式，通过微信公众号、微信转发、邮箱、社区网页广告等形式分发问卷。问卷中包括通用问题 22 项，针对不同领域的题目 17 项，题目如下（不含选项）：

1. 年龄范围
2. 性别
3. 最高学历（含在读）
4. 最高学历院校名称（全称）
5. 学习 AI 知识与技能的渠道（多选）
6. 获得 AI 信息的渠道（多选）
7. 希望从 AI 社区中获得的资源（多选）
8. 在 AI 社区中关注的功能
9. 所在领域属于（多选）
10. 常用编程语言
11. 你的 AI 研究方向
12. 常用 AI 开发工具/框架/平台
13. 常用的机器学习算法（非深度学习）
14. 常用的深度学习算法
15. 常用的 CV 模型架构（多选）

16. 常用的 NLP 模型架构（多选）
17. 感兴趣的预训练领域研究方向 你需要哪些支持资源
18. 2021 年你认为最值得关注的 AI 热点
19. 你对 2022 年中国 AI 发展的期待
20. 请用一句话形容 2021 年 AI 发展情况（非必填）
21. 请用一句话，概括你的研究或职业目标
22. 有效邮箱

对应不同领域的专属问题共 17 项：

◆ 如果所在领域是学术界：

1. 学术界身份
2. 科研团队规模
3. 每周平均阅读论文篇数
4. 每周科研工作时长
5. 每周教学工作时长
6. 哪些因素制约你做出好科研

◆ 如果所在领域是工业界：

1. 职位
2. AI 研发年限
3. 月薪资范围
4. 周工作时长

5. 所在行业
6. 工作单位性质
7. AI 团队规模
8. 融资轮次
9. AI 在贵单位中的使用情况

◆ 如果所在领域是投资界：

1. 如果创业，你看好
2. 如果投资，你看好

（二）数据清洗和分析方法

本调查采用基础的数据清洗和过滤规则。在分析有效问卷方面，我们采用了样本有效性筛选方法，包括重复填写问卷排除法、答题逻辑矛盾排除法、不完整问卷排除法等，过滤了最高学历院校无效、邮箱无效、回复选项随机填写、填写回答自相矛盾等情况。

数据分析方面，采用 Python 编写清洗和数据处理程序，Pandas 和 Numpy 进行数据分析和计算，采用 MongoDB 存储数据，采用 Bokeh 和 Pyecharts 进行数据可视化，形成柱状图、饼图等。对于选项为多选题的题目，仅生成柱形图，每一选项在受访者中的占比采用表格形式展示。对于选项较多的题目，仅展示 Top5 的结果。

(三) 调研存在的局限性

必须承认，本次调查存在方法上的局限。一是在样本收集方面，由于采用红包裂变等分发手段收集回复，可能导致问卷结果和社区实际情况有偏差。二是在数据处理方面，受制于人力、技术等资源，部分无效回复可能无法被及时发现。我们尽最大的可能进行筛查，确保结果可靠。

本次调查为智源社区首次大规模公开调查，欢迎广大用户提出意见反馈，我们将在后续的工作中不断精进，为用户提供更优质的内容和资源。此外，本次调查结束后，我们将对数据进行妥善处理，保护受访者的隐私和信息安全。

免责声明

本报告所含内容为一般性信息，不构成任何专业业务和财务业务的判断依据，同时本报告的信息来源于问卷调查结果，我们对该等信息的准确性、完整性或可靠性作尽可能的追求，但无法做任何保证和承诺，团队成员及所在单位不对任何因此报告导致的直接或间接损失或损害承担责任。

联系人：戴一鸣|智源社区分析师

邮 箱：ymdai@baai.ac.cn

■ 联系我们

电话: 010-68933383

邮箱: press@baai.ac.cn

网站: hub.baai.ac.cn



智源社区公众号