

# Datawhale

## 如何用开源方式构建学习者生态

苏鹏 Datawhale上海负责人



# Content 目录

- 01 Datawhale 介绍
- 02 Datawhale 文化
- 03 开源学习社区构建
- 04 未来展望



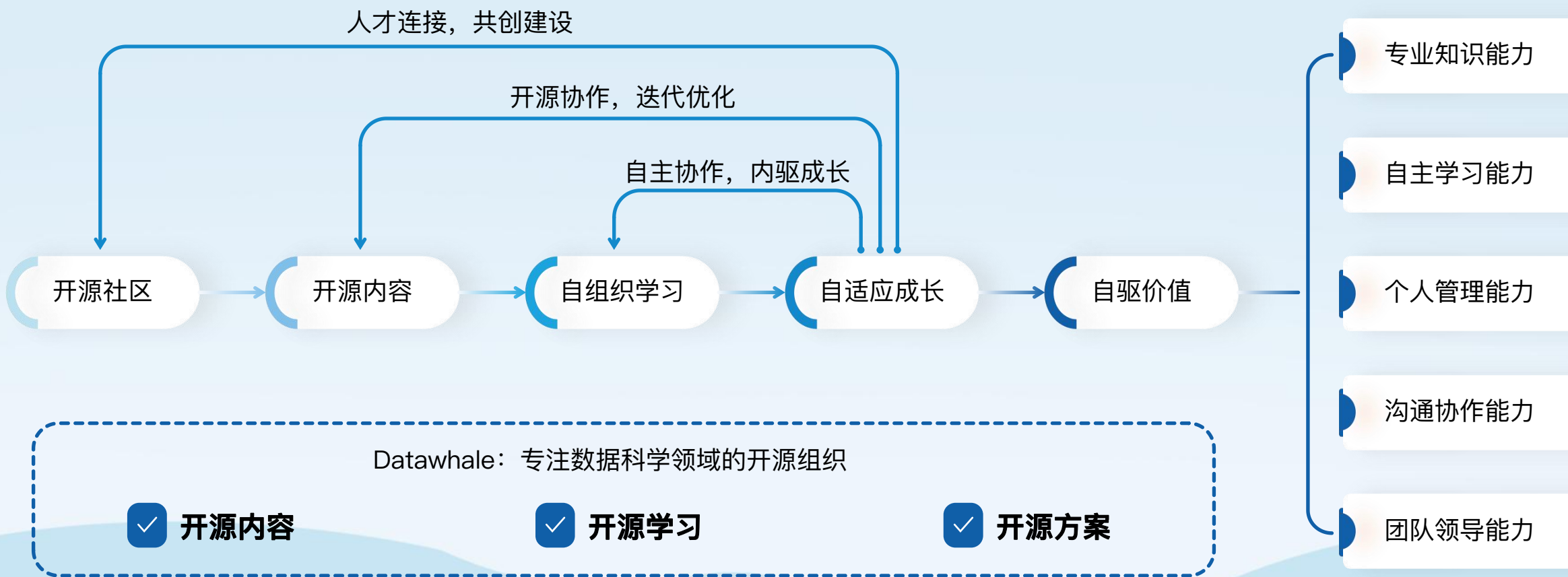
# Part 01

## Datawhale 介绍



# 开源学习社区 Datawhale

Datawhale 是国内首个开源学习社区，专注于数据科学和 AI 领域，  
以 for the learner 为使命，和学习者一起成长。



# 开源学习社区 Datawhale



清华大学、北京大学、中国科学院大学、浙江大学、上海交通大学、厦门大学、西安电子科技大学、华中科技大学、电子科技大学、西安交通大学、武汉大学、北京邮电大学、东北大学、华东师范大学、南京大学、同济大学、复旦大学、新加坡国立大学、卡内基梅隆大学、牛津大学、哈佛大学等

高校：2000+

贡献者：700+

生态伙伴：20+

阿里、百度、华为、科大讯飞、英伟达、蚂蚁集团、和鲸科技、Graviti、AIWIN、DataFountain、图森未来等

学习者：50W+

汇聚了来自伯克利、斯坦福、港中文、港科大、清华、北大、上交、复旦、浙大等全球近百所高校贡献者，以及微软、亚马逊、阿里、百度、腾讯、字节等在职的贡献者

**Datawhale 通过开源学习，帮助学习者 50w+ 人，覆盖全球 2000+ 高校。**



# 开源学习社区 Datawhale



策划主办/承办论坛、赛事、图书展、开源集市等板块，参与 GAIDC（全球人工智能开发者先锋大会）、WAIC（世界人工智能大会）等世界级盛会，促进人才发展，推动产业和人才的双向匹配。



GAIDC 中 AI 人才成长论坛，邀请孙茂松，俞勇，邱锡鹏等学术大咖，围绕人才成长分享探讨。



WAIC 滴水湖 AI 夜话，主办 AI 开源学习与开发者成长论坛，与 LF AI& DATA 基金会、Linux 基金会等开源组织与临港區达成合作。



2050 大会青年团聚与王坚院士

# 开源学习社区 Datawhale



2023年3月，Datawhale团队联合李沐老师，组织了动手学深度学习课程的学习。共吸引**9027人**参与学习交流，来自全球**733所高校**。



联合主办：OpenI启智社区，人民邮电出版社异步社区

特约合作伙伴：DataFun

赛事合作平台：阿里云天池

AI训练合作平台：趋动云科技

# 开源学习社区 Datawhale



2023年4月，Datawhale 联合 Hugging Face，吸引 **10425 人** 参加学习，同时邀请政产学研用各界大咖，围绕 AIGC 与大模型，举办 **8 场** 大咖讲座，**6W+** 人观看。

主办单位:

联合主办单位:

战略社区合作伙伴:

学习共创伙伴:

学习生态伙伴:

学习支持:

开源学习高校联盟

分享主题	分享嘉宾	嘉宾简介
引爆：ChatGPT/GPT-4背后的逻辑与应用趋势	Darren	工业和信息化部通信经济专家委员会办公室原副主任， 腾讯安全战略研究中心新技术安全首席研究员， 《AIGC未来已来》作者
GPT原理与LLM的挑战与机遇	周明	澜舟科技创始人兼CEO、 中国计算机学会 CCF 副理事长、 NLP和大模型领域的顶尖大牛
大模型下的算力挑战与机遇	Johny	趋动云总架构师
大模型时代的数据变革	李薇	上海人工智能实验室OpenDataLab产品主管
商业银行大模型（LLM）的建设与应用	潘淳	微软技术俱乐部（苏州）执行主席， 邮储银行苏州市分行“RPA+AI”创新实验室负责人
AIGC下的创业机遇与挑战	王建硕	百姓 AI 创始人&CEO
AIGC下的人才培养	金耀辉	上海交通大学电信学院教授、 上海交通大学人工智能研究院总工程师、 上海白玉兰开源开放研究院执行院长
大模型的伦理与关于Hugging Face的一切	王铁震	HuggingFace 中国区负责人、 前 TensorFlow 团队成员



# Part 02

## Datawhale 文化



## Datawhale 价值观 —— 保姆文化

### 社区第一

社区的整体利益和发展是放在首位的。所有的活动、项目和决策都以促进社区的和谐、进步和持续发展为目标。

### 以学习者为中心

Datawhale 将学习者的需求和成长放在核心位置。在提供课程内容、组织活动或制定学习计划时，都会首先考虑学习者的需求、学习风格和成长路径。

### 成就贡献者

强调对那些为社区做出贡献的个人或团队的认可和奖励。无论是通过分享知识、提供资源、参与项目还是以其他方式为社区贡献力量的人都应得到嘉奖和鼓励。

### 成为伯乐

每一个社区成员都可以成为“伯乐”，都有责任和能力去发掘潜在的人才，支持他们的成长，帮助他们实现自己的潜能。



# 开源学习社区 Datawhale

- 输出，一种底层的学习模式
- 组队，一种学习的场景
- 共同目标，驱动双方学习
- 找到工作，提供了正反馈

**“最早的那帮人，他们是很纯粹的，种子就播在这儿了...”**

“那时，另一个[同学]和我一样在实验室准备[找工作]...我说，我们现在这样学，等到学完,找工作就来不及了...要不我们两个一起来学习。”

“[我们]怎么学呢？我们可以先看一下招聘那些岗位缺什么东西，然后就发现缺项目...还需要算法...然后我就建议我们把最经典的十个算法拎出来，你讲给我听，我讲给你听...[我就发现]我大概[逐渐]能讲清楚了，反而学的更快了。然后缺项目，我就找了KDD Cup（竞赛项目），他也找了一个项目...后来反馈很明显...两个人都找到了相关的实习...后来[又]找到了算法工程师的工作...”

“找实习的时间，可能本来要三个月，[缩短到]一个月不到。”



# 开源学习社区 Datawhale



那些默默做事的人也是会被看到的。

第一年会的时候，大家在一起谈心，二三十个人，谈的是我其实是一个非常自卑的人，大家在谈自己不自信的那一面，自己心里面非常那什么的一面。

我理解那一批人，他们有一个非常强的特质，一个是比较谦虚，第二个是非常向上，即使是对一个非常有挑战性的任务，也想尝试一下，那时候，成员很多都是基于这种想法在做贡献。很多都是想，我自己学的还不够好，我想反向 push 我来学习。

——马晶敏

那一段时期的生活，我觉得是很纯粹的求知，跟其他人去讨论你的想法。大家会辩论，但是辩论的目的不是为了征服，是为了得到一个相同的结论。我想大家能够聚集起来也是因为找到了同类的人。

——李福

“大家其实会讨论，总有一个地方大家需要一个共识。在那个共识的时候谁都压根不敢保证说一定就对。但是在分歧的时候大家会吵，但是分歧之后一旦定了，大家就会按照这个去做，做完之后没有做成，没有人去斥责，他会让你很珍惜，你希望下一次做得更好...他会让我觉得我想要为这个组织在多付出一点点。”

——范晶晶



# Part 03

## 开源学习社区构建



# 开源学习社区 Datawhale

Datawhale 是国内首个开源学习社区，专注于数据科学和 AI 领域，  
以 for the learner 为使命，和学习者一起成长。

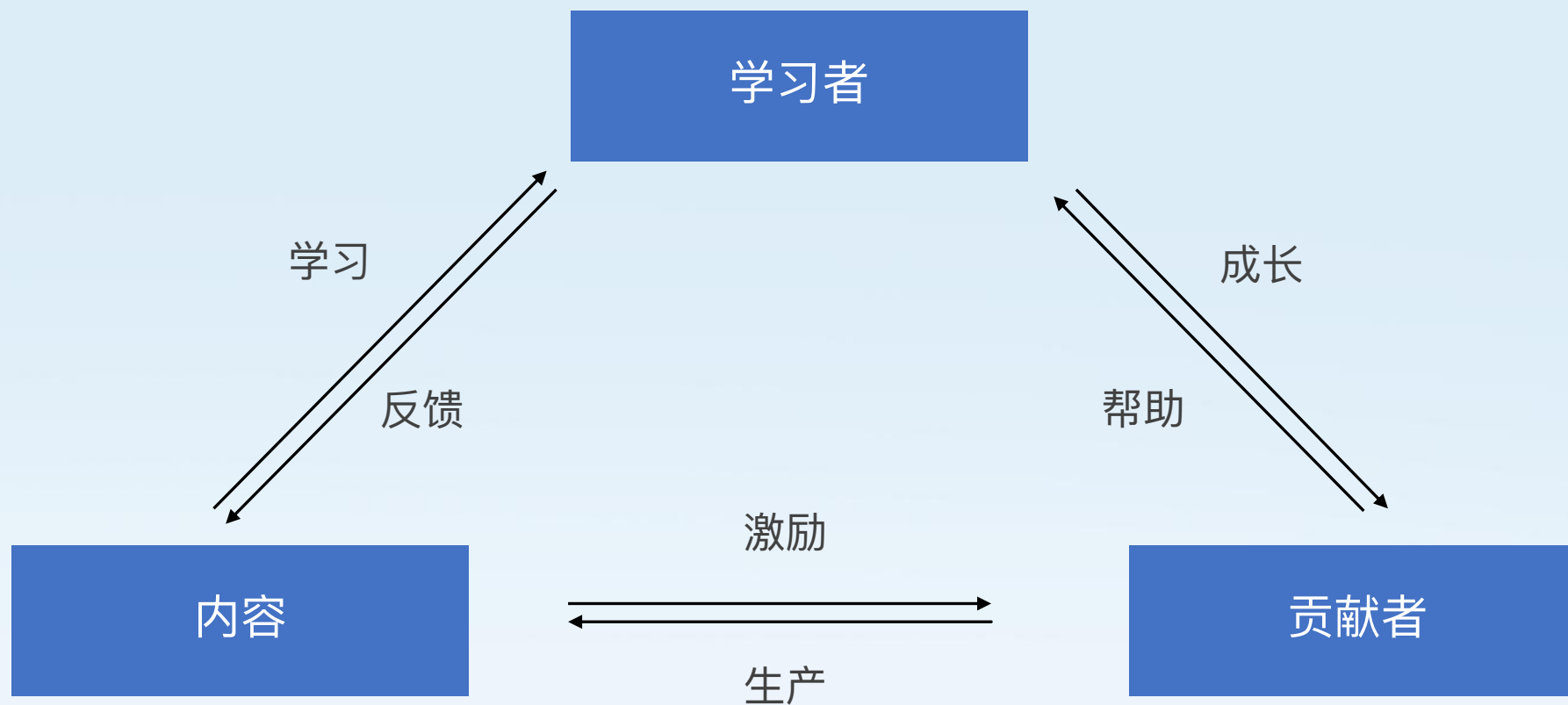


# 开源学习社区 Datawhale

开源学习（Open Collaborative Learning）是一种在开放环境下的协同学习模式。它强调学习者之间的互动、共享和合作，来共同创造知识、协同解决问题和创新应用。锻炼人才更符合产业需求的多维能力，建立新的评价体系。



# 开源学习社区 Datawhale





# 开源学习社区 Datawhale

出版的“南瓜书”、“蘑菇书”等书籍出版即为细分领域排名 Top，与人民邮电出版社，机械工业出版社共建数字化开源出版新模式。



Datawhale 开源了 70+ AI/数据科学领域的开源内容，构建了 [Datawhale人工智能培养方案](#)，从基础学习延伸至学术深耕、项目、竞赛实践，持续为人才成长赋能。

## Datawhale人工智能培养方案

- 深度学习公共必修课
  - 1、深度学习理论：水很深的深度学习
  - 2、深度学习理论：李宏毅机器学习笔记
  - 3、深度学习实践：深入浅出PyTorch
- 机器学习公共必修课
  - 1、机器学习理论：西瓜教程
  - 2、机器学习实践：西瓜书代码实战
- Python公共必修课
  - 1、Python基础：聪明办法学Python
  - 2、Pandas基础：Pandas数据处理与分析
  - 3、Numpy基础：巨硬的NumPy
- 数据分析方向
  - 必修
    - 1、SQL基础：奇妙的SQL
    - 2、Excel基础：自由Excel
    - 3、数据分析理论&实践：动手学数据分析
  - 选修
    - 1、集成学习理论&实践：ensemble-learning
    - 2、数据可视化：matplotlib奇遇记
    - 3、数据可视化：极好的plotly
  - 实践
    - 1、分类：零基础入门金融风控-贷款违约预测
    - 2、分类：零基础入门数据挖掘-心路信号分类预测
    - 3、回归：赛题介绍：根据汽车类型等信息预测二手汽车的交易价格
    - 4、综合：零基础入门数据分析之学术前沿趋势分析
    - 5、综合：数据竞赛Baseline & Topline分享
- 推荐系统方向
  - 必修
    - 1、推荐系统理论&实践：有趣的推荐算法
  - 选修
    - 1、集成学习理论&实践：集成学习
    - 2、强化学习理论&实践：强化学习教程
  - 实践
    - 1、综合：零基础入门推荐系统 - 新闻推荐
    - 2、综合：数据竞赛Baseline & Topline分享
- 计算机视觉方向
  - 必修
    - 1、计算机视觉理论&实践：OpenCV
    - 2、计算机视觉理论&实践：动手学CV
  - 实践
    - 1、语义分割：零基础入门语义分割-地表建筑物识别
    - 2、OCR：零基础入门CV - 街景字符编码识别
    - 3、综合：数据竞赛Baseline & Topline分享
- 自然语言处理方向
  - 必修
    - 1、自然语言处理理论&实践：NLP教程
  - 选修
    - 1、transformers库实践：基于transformers的自然语言处理(NLP)入门
  - 实践
    - 1、文本分类：零基础入门NLP - 新闻文本分类
    - 2、综合：数据竞赛Baseline & Topline分享

# 开源学习社区 Datawhale

**主题：**

**AI+X：**

希望将人工智能（AI）与各个学科和行业（X）结合，激发无限潜力和创造力（X），让年轻人拥有更多可能性（X）

**高校行：**

把 AI 带到 10 座城市，100 所高校，影响 10000 名高校师生

## AI+X 高校行



10座城市



100所高校



10000名师生

**动手实践工作坊，主题分享，学习交流，能力认证等实用好玩的 AI 线下活动，激发灵感和创新**

- AI+X 高校行活动融合人工智能企业最新技术与应用场景、Datawhale 丰富的学习资源、以及高校创新火种，打造独特的 AI 学习体验，激发高校学生的学习热情和应用能力，促进 AI 与各高校专业交融。
- 我们倡导学习者追寻兴趣热爱，包容开放，协作学习，动手实践，鼓励探索和创新。



# 开源学习社区 Datawhale



浙江大学



杭州电子科技大学



华东理工大学



同济大学



浙江工业大学



南京大学



南京邮电大学



上海交通大学

# Part 04

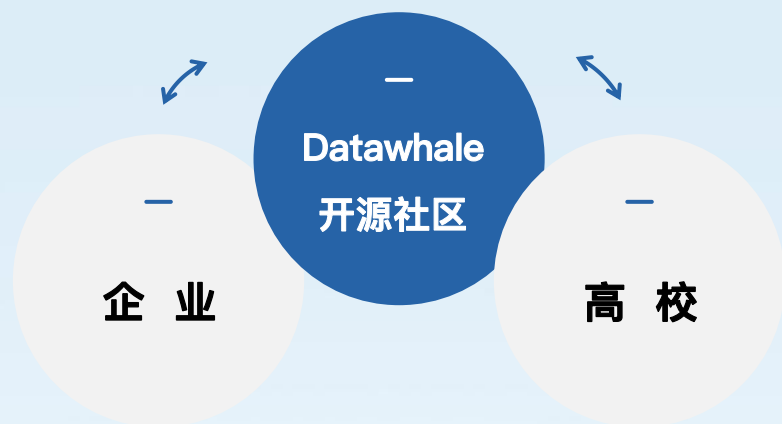
## 未来展望





# 开源学习社区 Datawhale

助力高校人才成长，联结企业人才需求



Datawhale 与企业合作共赢，与高校共同验证协作学习的模式，将人才成长诉求和未来就业、科研相连接，助力人才成长，建立起人与人、人与企业和人与未来的联结。

## Datawhale

## 社区赋能

核心团队

社区项目 Owner 20+

社区核心成员 300+

社区成员

社群成员 15W+

社区粉丝 50W+

创新孵化

课程创作

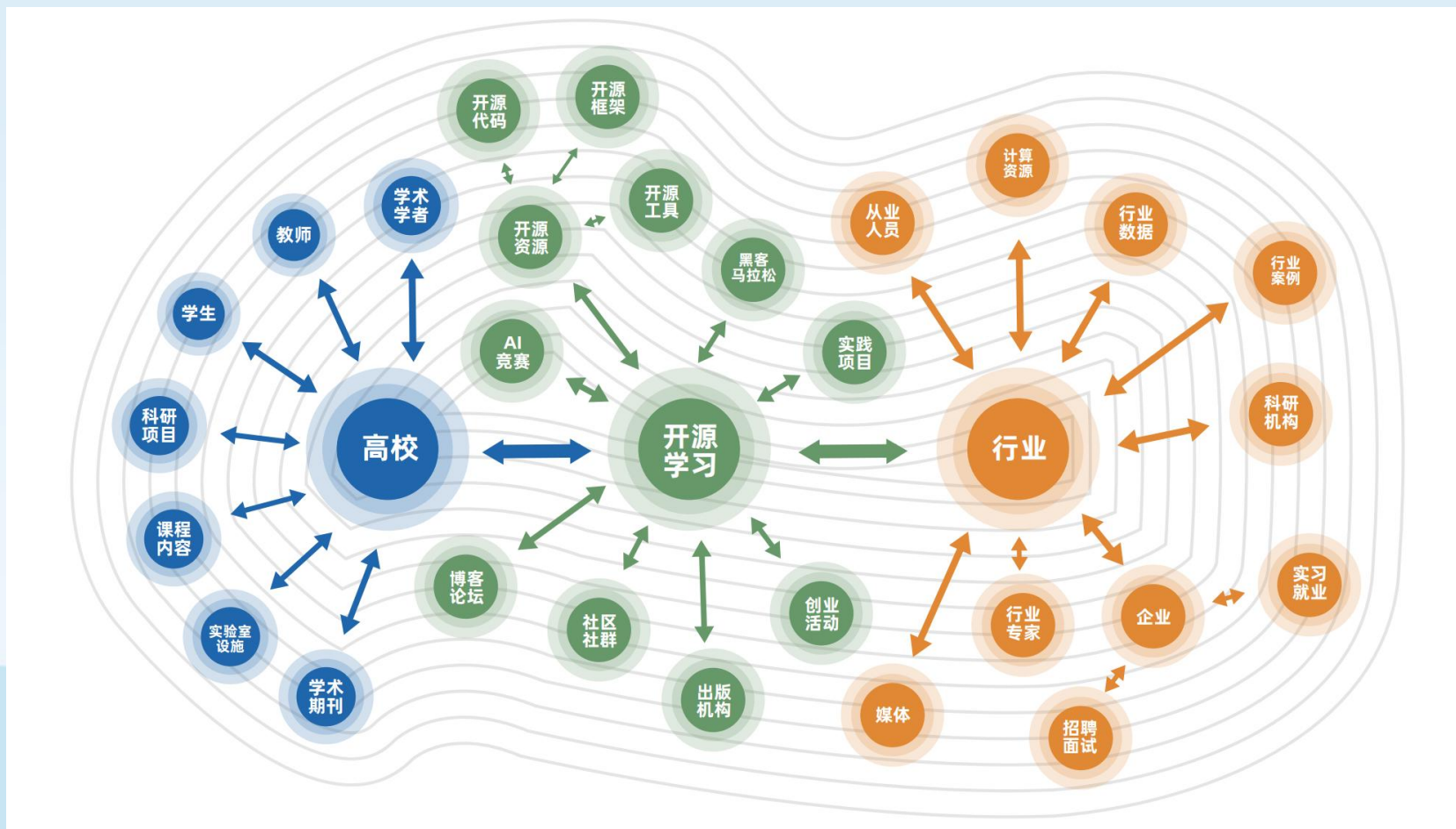
精准用户

社区生态

教育赋能

# 开源学习社区 Datawhale

开源学习构建了多元学习生态网络，让人才无限接近行业： 开源学习构建了一个链接高校和产业的网络，网络上的二级元素可以自由链接进行组合，形成特定的学习场景。让人才无限接近行业，把人才能力与场景需求、相关资源结合起来，实现创新和应用价值。





# Thanks.

