第5届全国高校大数据与人工智能教学研讨会

2022.05.13–2022.05.14　中国·厦门

大会官网

主办单位：教育部高等学校计算机类专业教学指导委员会

承办单位：厦门大学 XIAMEN UNIVERSITY　天津大学 Tianjin University　中南大学 CENTRAL SOUTH UNIVERSITY　西南财经大学 SOUTHWESTERN UNIVERSITY OF FINANCE AND ECONOMICS　福州大学 FUZHOU UNIVERSITY　华侨大学 HUAQIAO UNIVERSITY　集美大学　福建技术师范学院　人民邮电出版社 POSTS & TELECOM PRESS

协办单位：美林数据 MERITDATA　TRANSWARP 星环科技　海豚实验室　山东省计算中心 SHANDONG COMPUTER SCIENCE CENTER　荔峰科技 LIFENG　Bai du 百度　商迪3D web3di.com

海豚实验室

大数据与人工智能专业与实验室建设“新范式”

吴仲毓
杭州睿数科技　创始人& CEO

公司简介

海豚实验室

杭州睿数科技有限公司

- 坐落在杭州硅谷-未来科技城CBD核心区欧美金融城。是一家由海归人才创办的高新技术公司。创始团队由来自IBM、Facebook、阿里云等世界顶尖IT企业高管和专家组成。

- 公司秉承“*赋能教育、赋能人才*"的理念，自研了“海豚人工智能与大数据实验室整体解决方案，已经获得国内众多高校的广泛认可，并相继与国内几十所“双一流”、“双高计划”的高校展开合作，并实现项目成功落地。

理念/价值观

赋能教育　赋能人才

产品/服务

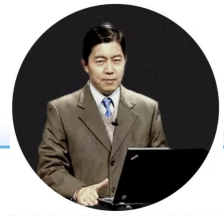
云实验室
大数据/人工智能实验室
培训与教育服务

企业愿景

人工智能与大数据教育
整体解决方案的领导者

特聘高校专家名师

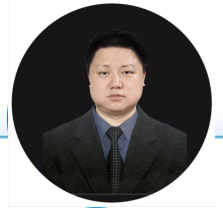
海豚实验室



赵卫东

复旦大学 教授

主要研究方向包括机器学习应用和大数据分析等。主持国家自然科学基金2项、国家重点研发计划子课题、上海市浦江人才以及企业合作课题等30多项目。已在国内外刊物和学术会议发表论文100多篇。出版专著、教材《机器学习》《机器学习案例实战》《Python机器学习实战案例》等10多部。获得上海市2015年上海市科技进步二等奖。CDA三级认证数据科学家，腾讯云和百度云机器学习认证讲师。



陈昊鹏

上海交通大学 教授

中国计算机学会高级会员。于2001年获得西北工业大学计算机科学与工程系博士学位，2002年3月进入上海交通大学计算机科学与技术博士后流动站工作，2004年3月博士后出站在上海交通大学软件学院任教至今。 他一直从事分布式计算方面的教育和研究工作。也曾承担过863专题课题负责人，并获得发明专利3项、软件著作权3项， 并参与教育部，上海市经信委等多项产学研项目，在国际会议或期刊上发表相关学术论文50余篇。

王万良

浙江工业大学 教授

原浙江工业大学计算机学院院长，教授，博士生导师，国家万人计划领军人才，享受国务院政府特殊津贴专家，国家级教学名师，浙江省杰出教师。现任教育部高等学校计算机类专业教学指导委员会委员、全国高校大数据教育联盟副理事长、浙江省高等学校计算机类专业教学指导委员会副主任、浙江省计算机学会副理事长、浙江省计算机应用与教育学会副理事长、杭州市政府参事、杭州市科协常委、

CONTENTS

目录

1 数智化人才 *"新需求"*

2 专业建设 *"新范式"*

3 人才培养 *"新模式"*

4 海豚实验室 *"新标准"*

数智化人才 “新需求”

PART 01

企业数字化转型要具备新的数字化领导力

海豚实验室

- 企业的数字化转型一定需要从企业的顶层设计开始。所以，企业的数字化转型必须要首先从企业的最高管理者开始。企业最高管理者要首先转型。企业的主要管理者必须要基本弄清楚未来数字化发展的方向，转换新的数字化管理企业理念，学习掌握一定的数字化基础知识。

企业运营逻辑

信息化

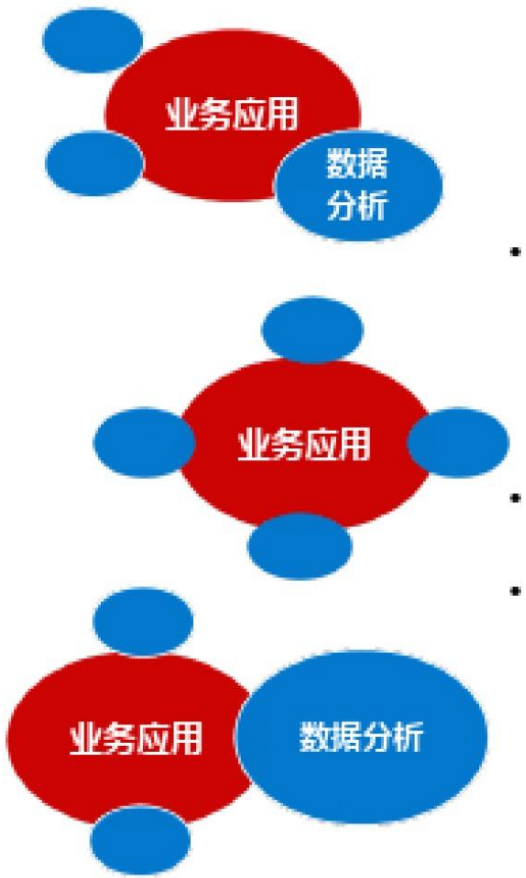
信息化系统把企业内部的运行用这套信息系统全部管理规范起来。并且甩掉了企业以往可以看得见的帐、单、表、证。解决了企业内部管理的流程效率问题。信息化是用系统去管理人、流程、工作。

数字化

借助数字化能力打通企业内部与外部、打通企业各个环节之间的链接。以连接为主线，重构企业的模式、流程。把企业的相关业务动作，迁移到网络环境下完成，使企业的运行效率获得显著的提升。

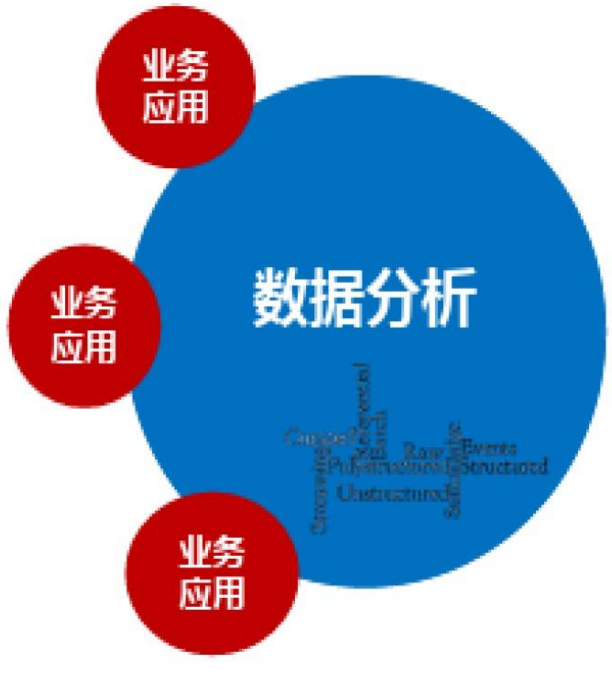
- 企业数字化变革，并且在整个的转型过程中，必须是要**最高管理者直接规划、组织、指挥转型**。管理者没有搞清楚不要转，管理者还没有转变理念不能转。**最高管理者首先具备数字化的领导力，是企业转型的基础**

从以应用为中心



- 基于业务应用的数据分析(仅分析来自于本业务应用的数据）
- 面向应用的分析人员
- 面向应用的数据管控

到以数据分析为中心



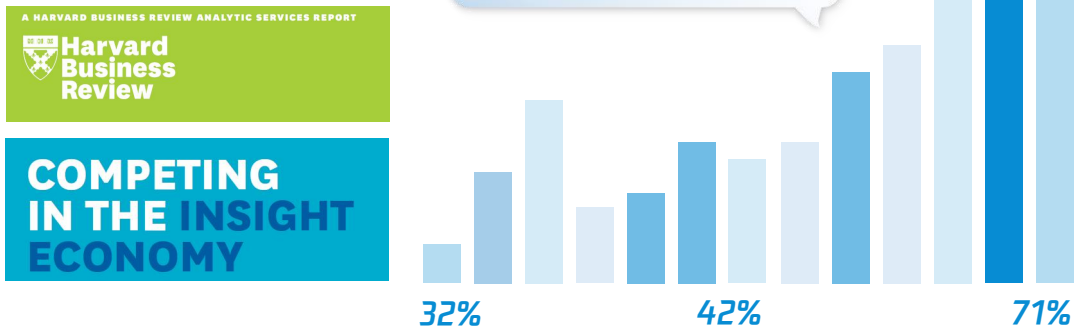
- 综合性数据分析平台
- 专业的数据分析团队
- 统一的数据管控机制

驱动力：云

- 业务应用的数据资源统一集中到云

数据科学支撑的大数据与人工智能

海豚实验室



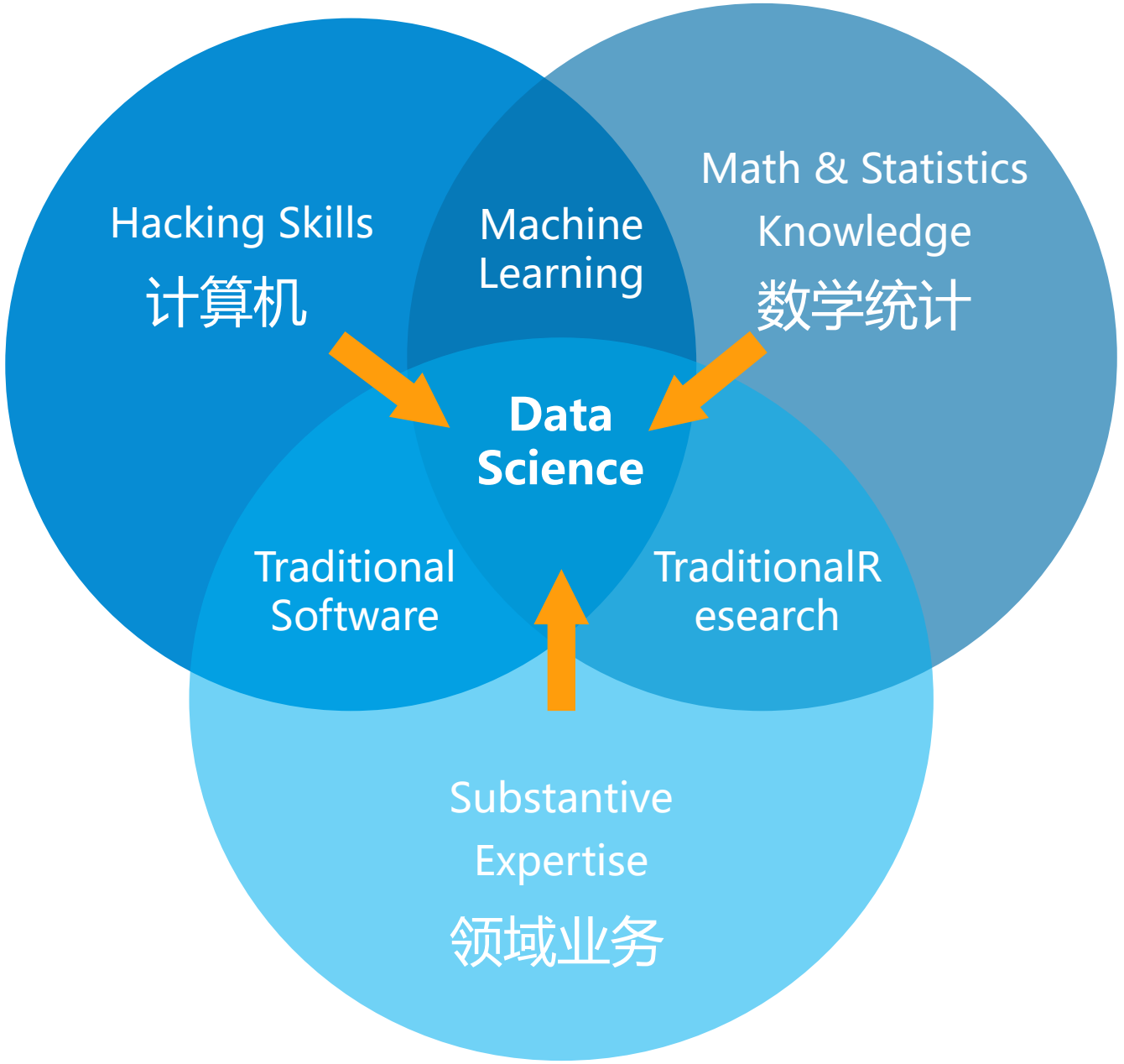
71% 大数据用户从洞察中创造新的经济价值和商业机会

42% 大数据用户称未来三年数据驱动的洞察将成为未来业务的增长点,发挥更大的作用

32% 大数据用户将数据和洞察变现

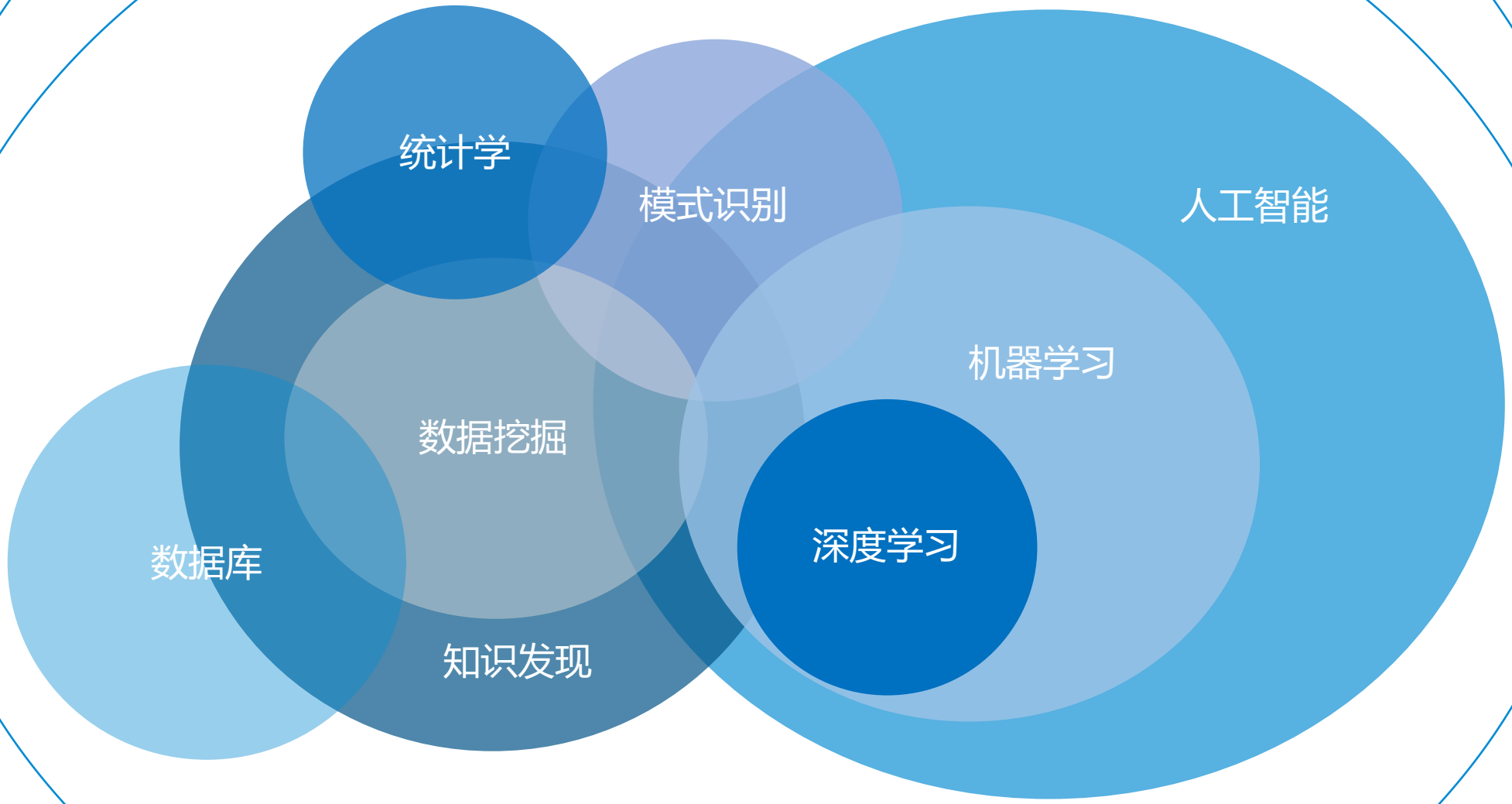
数据科学专业群人才培养核心

- “全”数据
- 数据分析方法
- 数据驱动的思维
- 数据工程能力


Hacking Skills
计算机
Machine Learning
Math & Statistics Knowledge
数学统计
Data Science
Traditional Software
TraditionalResearch
Substantive Expertise
领域业务

AI 知识图谱

海豚实验室



什么是数据科学？

海豚实验室

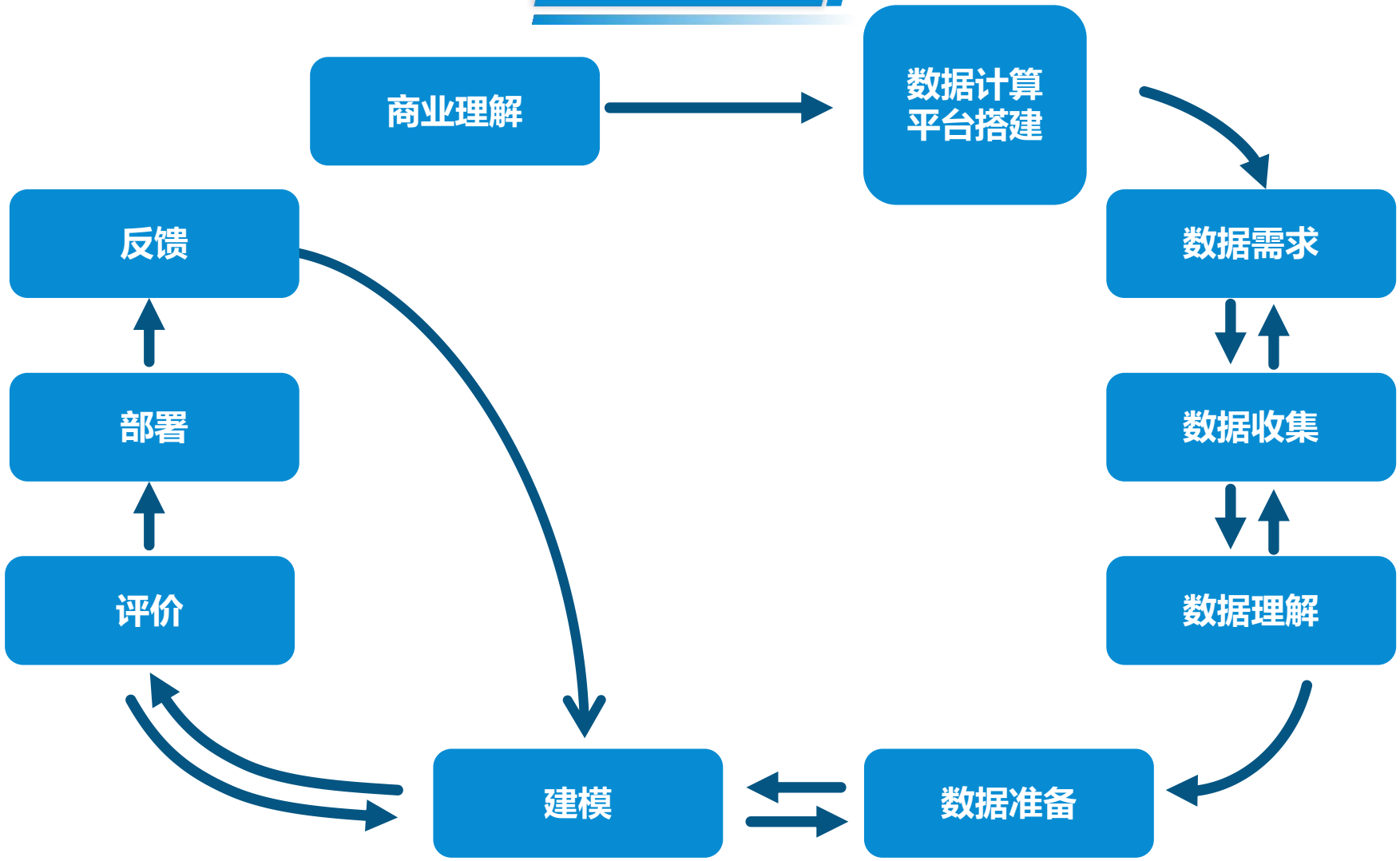
将数据转化为洞见

数据科学是一个从数据(结构化或非结构化的大数据)中提取知识或洞见的跨学科的领域，是某些数据分析如数据挖掘等的延续。

数据科学方法论

海豚实验室

方法论图解



新基建背景下的传统行业数智化升级，人才需求非常大

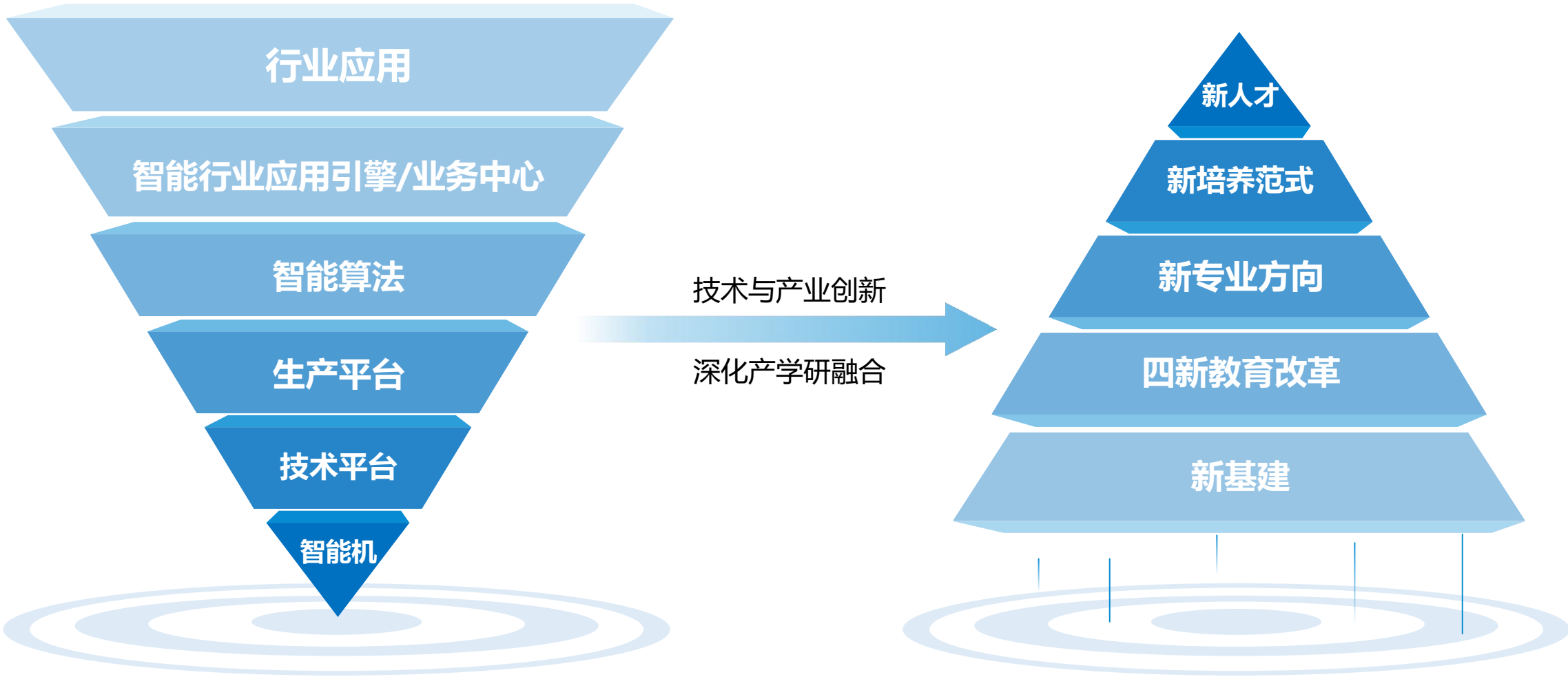
- 医疗 - 我如何改善医疗诊断？
- 商业 - 我如何增加销量？如何展开个性化的交流
- 交通 - 交通信号灯应该亮多长时间？
- 心理学/神经系统科学 - 我们可以预测人的思维和行为吗？
- 运动 - 我们可以预测一个运动队的表现吗？
- 社会公益 - 我们如何用数据来减少贫困，改善生活质量？
- 银行业 - 我们可以更准确地分辨出欺诈行为吗？
- 零售 - 如何使顾客购买更多我们的产品？
- 广告 - 如何使我们的广告更加的个性化？

大数据与人工智能专业建设
"新范式"

PART 02

融合产业链与教育链、人才链与创新链，探索数据科学人才培养模式

海豚实验室



产业用人需求
人才评估标准
海豚人才培养内容体系
人才岗位
行业规范
工程能力
专业背景
行业经验
实践能力
理论基础
海豚大数据/人工智能专业培养方案
教学资源
课程体系
实验资源
实训案例
行业数据集
实验实训
支撑平台
教学管理
考评竞赛
软/硬件环境
师资赋能
师资研修班
内容共建
科研共创
权威认证
终身学习者
科技领军人才
数据科学家
数据工程师
产业工人

专业融合

总体目标：培养AI全产业各层次人次

校企合作、共建培养方案、共建课程体系、执行教学任务

课程教学

精品课程 进度可控

智能互动 随时答疑

班级管理 企业实习

实验案例教学

实践实训 效果可控

行业案例 真实数据集

详细教案 配套实验

个人

团队

学生训练营

项目答辩

动手实践

理论授课

数智化人才培养和实践平台

培养方法融合 学生素能融合

人才培养“新范式”

海豚实验室

数据探索式的学习

高密度实验实训

实验案例教学

数智化人才岗位需求

海豚实验室

数据工程师

针对大数据储存查询需求，完成大数据系统架构设计以及研发工作

① 程序开发 (Python、Java...)
② 数据库设计 (MySQL、MongoDB...)
③ 分布式储存系统 (HDFS...)
④ 分布式计算系统 (Spark...)
⑤ 数据预处理 (矩阵转置，缺失值填充...)
⑥ 多维特征分析 (关联性分析，主成份分析...)
⑦ 熟悉数据软件和工具
⑧ 基础算法 (分类，聚类...)

数据分析师

根据明确业务分析预测目标完成数据采集，数据处理分析建模等技术工作

① 大数据底层原理
② 传统算法 (分类、回归、聚类...)
③ 时间序列分析
④ 数据可视化
⑤ 文本挖掘算法 (分词、文本分类)
⑥ 数据分析、建模
⑦ 业务建模理解
⑧ 行业知识

数据科学家

根据客户目前业务特点以及数据现状，规划构建人工智能分析应用场景

① 大数据底层原理
② 传统算法 (分类、回归、聚类...)
③ 时间序列分析
④ 数据可视化
⑤ 文本挖掘算法 (word2vec、语法摘要)
⑥ 深度学习 (神经网络...)
⑦ 商业模型理解
⑧ 专业的行业知识

人工智能岗位需求

海豚实验室

数据标注工程师

① 程序开发 (Python、Java...)
② 数据库设计 (MySQL、MongoDB...)
③ 分布式储存系统 (HDFS...)
④ 分布式计算系统 (Spark...)
⑤ 数据预处理 (矩阵转置，缺失值填充...)
⑥ 多维特征分析 (关联性分析，主成份分析...)
⑦ 熟悉数据软件和工具
⑧ 基础算法 (分类，聚类...)

人工智能测试工程师

① 机器学习算法评估指标
② 机器学习算法概论
③ 数据分析处理
④ 自动化测试脚本工具
⑤ 标签数据集构建管理
⑥ 数据可视化
⑦ 熟悉数据软件和工具
⑧ 分布式环境压力测试

机器学习工程师

① 机器学习算法原理
② 传统算法（分类、回归、聚类...）
③ 神经网络算法应用
④ 自然语言处理算法应用
⑤ 图形处理算法应用
⑥ 人工智能模型调优
⑦ 应用开发（Python/Java/SQL）
⑧ 业务场景实现

“数据科学与大数据技术”专业课程体系架构图

海豚实验室

知识提升	机器人与人机交互	智能制造	无人驾驶	商业智能技术
实践技术	推荐系统		海量数据分布式开发	
	数据仓库		用户画像特征分析系统	
	网络爬虫与数据收集		商业预测系统	
核心专业课	大数据实时流处理		机器学习原理与应用	
	开源大数据工程基础(Hadoop)		数据分析技术	
	MySQL数据库技术		数据可视化方法与实践	
	Linux操作系统/ Python程序设计基础 / Java面向对象程序设计			
基础理论	线性代数 / 概率论与数理统计 / 高等数学 / 离散数学 / 模拟电子技术 / 数字电子技术			
	模拟系统原理及应用 / 计算机组成原理 / 计算机网络 / 数据结构 / 数据库原理			

人工智能专业的知识结构及课程体系 (参考： 卡内基•梅隆大学 CMU)

海豚实验室

认知、能力综合层：重要度（1），共16+16+32=64学时，4学分

社会与机器伦理 <16>　人工智能前言讲座 <16>　面向应用的智能系统设计与开发 <32>

特色模块层：重要度（3），从本层同一模块（或含核心层、平台层）中选择不少于64学时，4学分

智能感知与人机交互　机器学习与知识工程　自然语言处理　智能机器人与无人系统　智能应用技术

智能感知与人机交互	机器学习与知识工程	自然语言处理	智能机器人与无人系统	智能应用技术
智能传感与监测 <32>	知识工程与专家系统 <32>	语言识别 <32>	智能传感与监测 <32>	智能可穿戴技术<32>
图像分析与理解 <32>	群体智能与分布智能 <32>	机器翻译 <32>	图像分析与理解 <32>	智能终端技术 <32>
智能信息获取与处理 <32>	集成学习与强化学习 <32>	自然语言生成 <32>	智能信息获取与处理 <32>	智能游戏 <32>
多模式感知与处理 <32>	知识图谱技术 <32>	人机对话交流 <32>	多模式感知与处理 <32>	……
人机协同交互 <32>	计算智能 <32>	情感计算 <32>	人机协同交互 <32>	智能医疗技术 <32>
脑机接口 <16>	智能决策支持系统 <16>	自动文摘 <16>	脑机接口 <16>	智能教育技术 <32>
人机混合增强智能 <16>	类脑智能与类脑学习<16>	文本检索与信息抽取 <16>	人机混合增强智能 <16>	智能金融技术 <32>

模块平台层：重要度（2），选择本层（或含核心层）中不少于32*2=64学时，4学分

机器学习（下）<32>， 神经网络与深度学习<32>， 自然语言处理<32>， 机器视觉<32>，智能机器人<32>

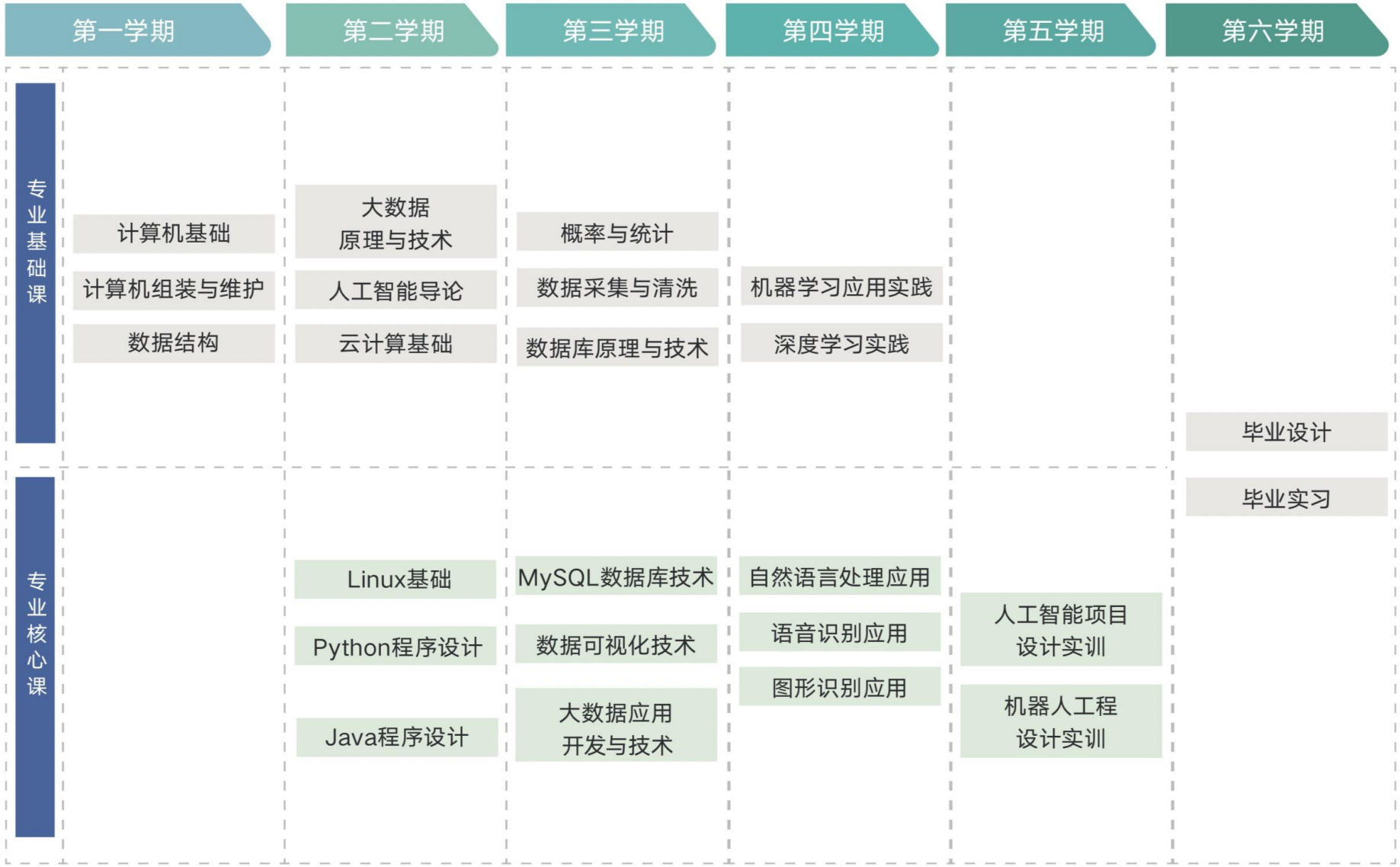
专业核心层：重要度（1），选择其中不少于 16+32*3=112学时，7学分

智能科学与技术导论<16>， 脑域认知基础<16>， 人工智能基础<32>， 机器视觉（上）<32>，数据挖掘<32> 模式识别<32>

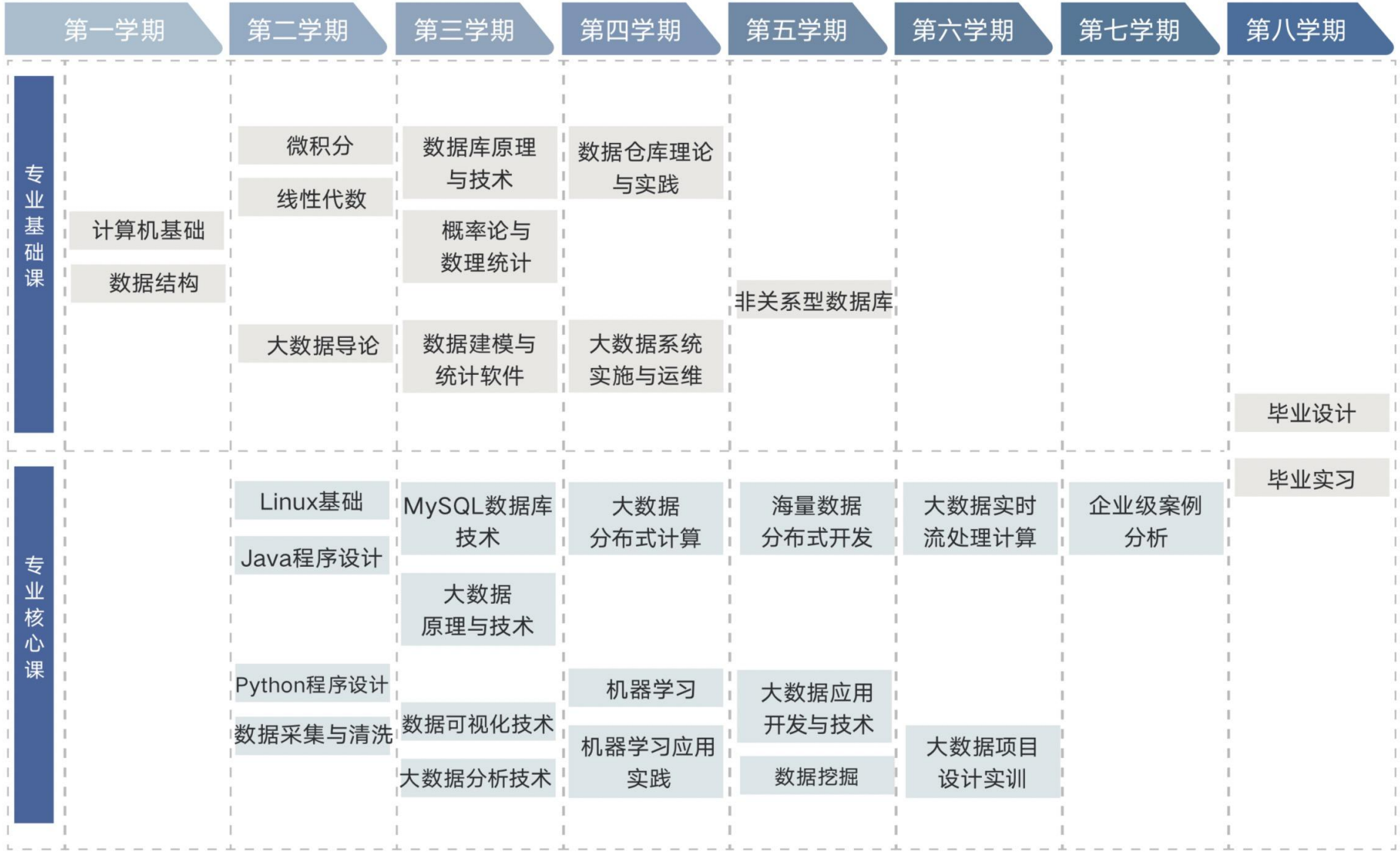
• 人工智能/智能科学（本科）

第一学期
第二学期
第三学期
第四学期
第五学期
第六学期
第七学期
第八学期
专业基础课
计算机基础
数据结构
微积分
线性代数
概率论与数理统计
数据库原理与技术
大数据原理与技术
人工智能导论
大数据分析
机器学习
嵌入式系统开发
深度学习基础
最优化理论与算法
语言模型
模式识别
毕业设计
毕业实习
专业核心课
Linux基础
Python程序设计
Java程序设计
数据采集与清洗
数据可视化技术
人工智能开源框架
自然语言处理
数字图像处理
文本分析与挖掘
图像识别技术
机器人操作系统
强化学习
迁移学习
语音识别技术
智能机器人设计与应用
人工智能项目设计实训
机器人工程设计实训

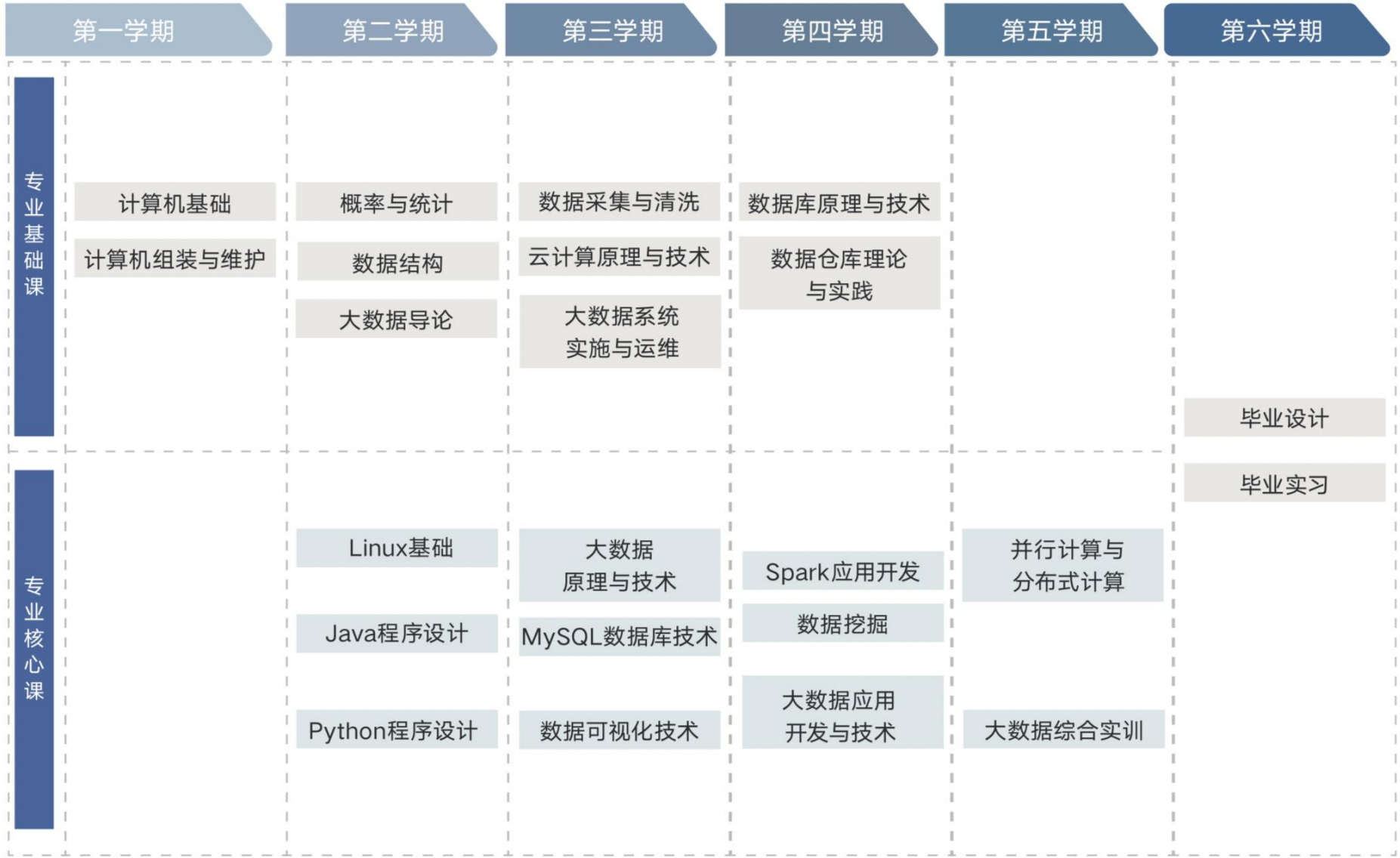
• 人工智能技术服务（高职）


第一学期
第二学期
第三学期
第四学期
第五学期
第六学期
专业基础课
计算机基础
计算机组装与维护
数据结构
大数据原理与技术
人工智能导论
云计算基础
概率与统计
数据采集与清洗
数据库原理与技术
机器学习应用实践
深度学习实践
毕业设计
毕业实习
专业核心课
Linux基础
Python程序设计
Java程序设计
MySQL数据库技术
数据可视化技术
大数据应用开发与技术
自然语言处理应用
语音识别应用
图形识别应用
人工智能项目设计实训
机器人工程设计实训

• 数据科学与大数据技术


第一学期
第二学期
第三学期
第四学期
第五学期
第六学期
第七学期
第八学期
专业基础课
计算机基础
数据结构
微积分
线性代数
大数据导论
数据库原理与技术
概率论与数理统计
数据建模与统计软件
数据仓库理论与实践
大数据系统实施与运维
非关系型数据库
毕业设计
毕业实习
专业核心课
Linux基础
Java程序设计
Python程序设计
数据采集与清洗
MySQL数据库技术
大数据原理与技术
数据可视化技术
大数据分析技术
大数据分布式计算
机器学习
机器学习应用实践
海量数据分布式开发
大数据应用开发与技术
数据挖掘
大数据实时流处理计算
大数据项目设计实训
企业级案例分析

• 大数据技术应用（专科）


第一学期
第二学期
第三学期
第四学期
第五学期
第六学期
专业基础课
计算机基础
计算机组装与维护
概率与统计
数据结构
大数据导论
数据采集与清洗
云计算原理与技术
大数据系统实施与运维
数据库原理与技术
数据仓库理论与实践
毕业设计
毕业实习
专业核心课
Linux基础
Java程序设计
Python程序设计
大数据原理与技术
MySQL数据库技术
数据可视化技术
Spark应用开发
数据挖掘
大数据应用开发与技术
并行计算与分布式计算
大数据综合实训

人才培养“新模式”

PART 03

2018年教育部授牌 ： 企业创新学生训练营

海豚实验室

Project-Driven Learning：项目驱动式学习

教育部高教司二级巡视员、理工处处长吴爱华（右一）为企业授牌

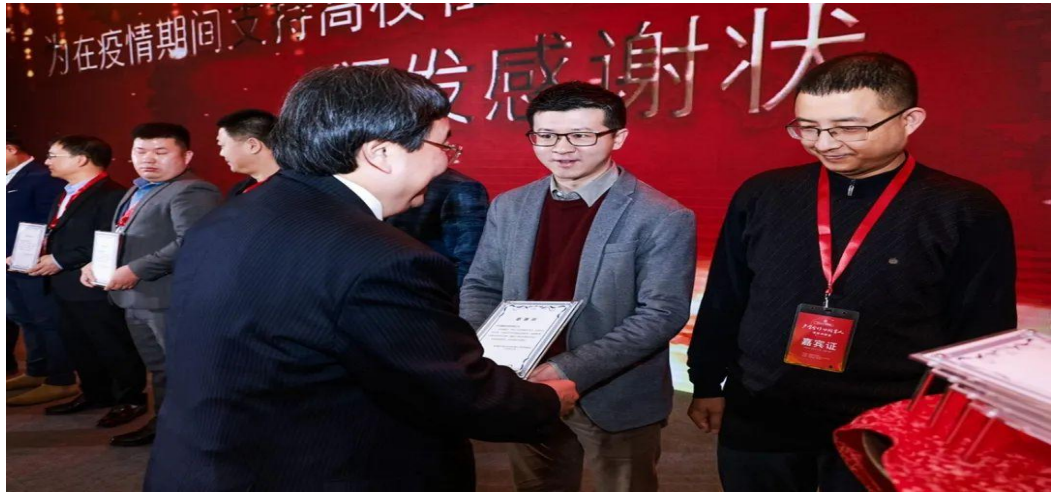
海豚实验室

教学大纲

数据科学导论与Python基础	数据科学简介、 海豚大数据实验室基本使用介绍、 Python 基础入门、 Python 基础实验
数据分析	数据科学的方法论和应用、 数据科学分析讲解、 数据分析 I 、 王者荣耀英雄特点分析、2012年欧洲杯数据分析、 北京房价数据分析 世界酒类饮品数据分析
数据统计	数据统计及方法理论简介、 数据统计的应用、 数据分析 II《柴油汽车耗油率》数据整理方法练习、 基于 Python 数据统计案例练习《波士顿住房》
数据可视化	数据可视化简介、 数据可视化工具的使用、 数据可视化简介 I《美国人的死亡原因是什么？》、 数据可视化简介 II《绘制中国天气数据的地图》
机器学习	机器学习的基础介绍、 机器学习常用算法的介绍、 机器学习算法 I、 机器学习算法 II、机器学习算法 III
项目答辩	学员分组，项目设计、开发、专家评审

2020年荣获教育部感谢状

海豚实验室



感 谢 状

杭州睿数科技有限公司：

疫情期间，贵公司积极响应号召，提供优质资源，为高校正常开展在线教学、保障教学质量做出突出贡献，展现了良好的责任担当！

特颁发感谢状，以资表彰与鼓励。

教育部产学合作协同育人项目专家组
2020年12月

教育部高教司副司长范海林（左一）为企业授感谢状

海豚实验室

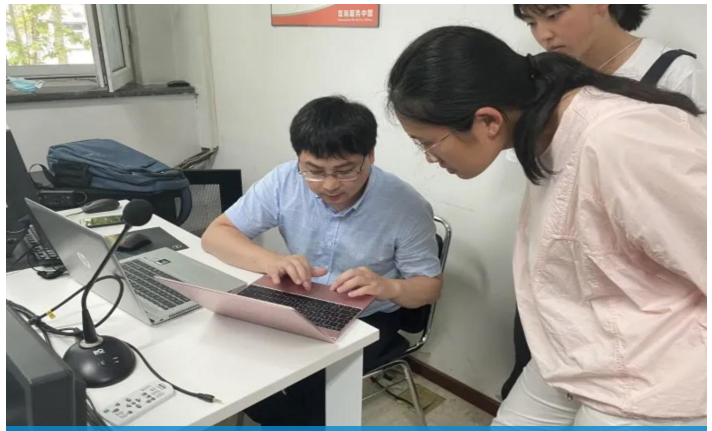
北京师范大学人工智能学院

6月28-7月2日 《数据科学与人工智能训练营项目答辩暨结业典礼》在北师大成功举办！

数据科学训练营 - 北京师范大学

海豚实验室

实训营现场



同学们积极准备数据科学项目

项目答辩现场（珠海校区）

答辩现场（北京校区）

学员背景：北京师范大学人工智能学院北京校区和珠海校区的人工智能、大数据和计算科学与技术专业，共计140多名本科生

实训营形式：为期一周的训练营学习，包括基础模块知识的学习，快速掌握人工智能相关的理论知识，技术和工具。

海豚实验室

《数据科学训练营》结业及成果展示

部分结业项目表
1 .意大利北部沿海城市气象数据可视化与分析
2. 视频弹幕可视化及基于Bi-LSTM和传统SVM模型的情感分析
3. 基于 YOLOV5的口罩佩戴智能检测模型
4. 基于LSTM模型的淘宝商品评价可视化分析
5. 简单的手势识别
6. 不同情感分析方式探索
7. 共享单车用户群体可视化分析
8.基于实时手势识别的拼图游戏
9. 基于卷积神经网络的新冠肺炎CT胸片分类
10. 基于卷积神经网络的手写数字识别

.....

海豚实验室

浙江大学控制科学与工程学院

本次《数据科学基础训练营》由 浙江大学控制科学与工程学院与杭州睿数科技共同举办，从全院众多报名学生中挑选出了50人，并于10月22日，成功举办了《海豚大数据竞赛暨结营仪式》。

数据科学训练营 - 浙江大学

海豚实验室

实训营现场

侯迪波副院长致辞开营

学生答辩现场 一

答辩现场 二

学员背景：全院众多报名学生中挑选出了50人，涉及自动化、电子信息等专业各年级的学生。

实训营形式：通过32个学时，包括“Python基础”，“数据分析”，“机器学习”，“数据可视化”、“深度学习”、“行业案例实战”六大模块的学习，

《数据科学训练营》结业及成果展示

部分结业项目表
1、贷款违约预测
2、二手车交易价格预测
3、大脑肿瘤CT影像分析
4、图书推荐系统
5、互联网情感分析
6、基于脱丁烷塔预测丁烷含量分析
7、 Airbnb房价预测
8、特朗普推特分析

.....

数据科学训练营 – 浙江万里学院

海豚实验室

2021年6月，由杭州睿数科技有限公司与浙江万里学院-大数据与软件学院联合打造的“数据科学训练营”迎来了54名结业学生

数据科学训练营 – 浙江万里学院

海豚实验室

《海豚数据科学训练营》

项目答辩现场（一）

项目答辩现场（二）

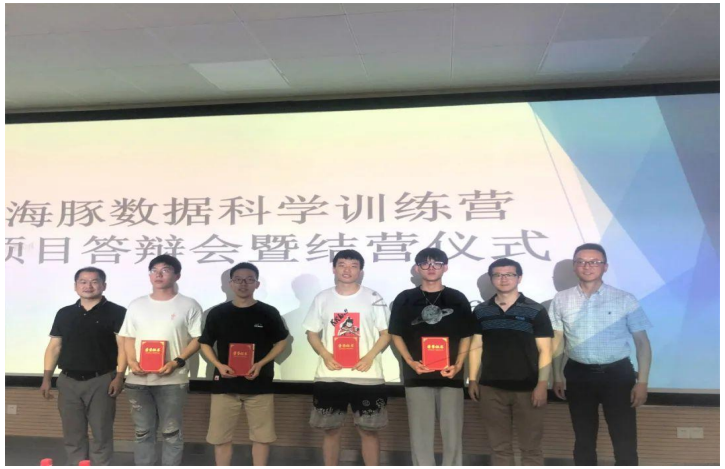
项目答辩现场（三）

学员背景：大数据与软件学院大二、大三的学生共54人

实 训 营： 1周的线下培训 + 3周的项目设计。课程以Python作为编程语言，分为数据科学基础、Python编程基础、统计分析、数据可视化、数据分析以及数据科学项目答辩六大模块，共计32个学时

《数据科学训练营》结业及成果展示

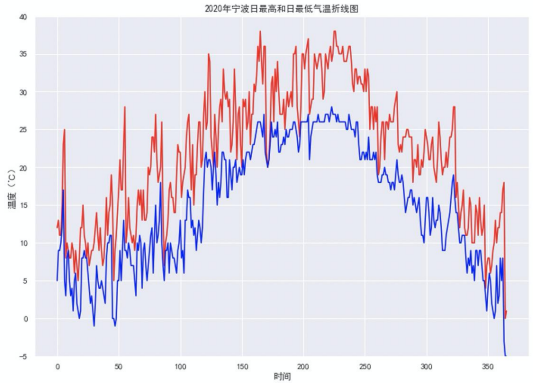
部分结业项目表
1. 疫苗接种情况分析与数据可视化
2. 电子游戏销售分析与可视化
3. 恐怖袭击数据分析与可视化
4. 淘宝粽子节日销售数据分析
5. 2020年宁波天气特点数据分析
6. NBA球员风格与变化的数据分析
7. 樱桃产地的产量与销售数据分析与预测
8. Python职业数据分析
9. 股票数据分析
10. 神奇宝贝属性相关性分析
11. 北京朝阳医院销售量分析



2020年宁波天气特点分析

结合上面两个图，可以绘制出2020年宁波日最高和日最低气温折线图

```
In [28]: bWendu=data['bWendu']
         yWendu=data['yWendu']
         plt.plot(bWendu,color='red',linestyle='-')
         plt.plot(yWendu,color='blue',linestyle='-')
         plt.xlabel('时间')
         plt.ylabel('温度（℃）')
         plt.ylim((-5.0,40.0))
         plt.title('2020年宁波日最高和日最低气温折线图')
         plt.show()
```

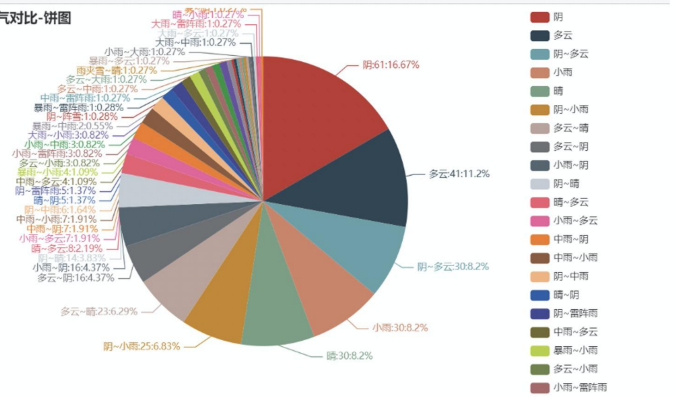


从绘制出来的2020年宁波日最高和最低气温折线图来看，该年宁波日最高气温和日最低气温走势基本一致，除七月下旬左右之后，昼夜温差稍大一些，其余时间昼夜温差都不是很大。

绘制宁波2020天气对比-饼图

```
In [34]: pie=create_pie(datas,"宁波2020天气对比-饼图")
         pie.render_notebook()
```

Out[34]:



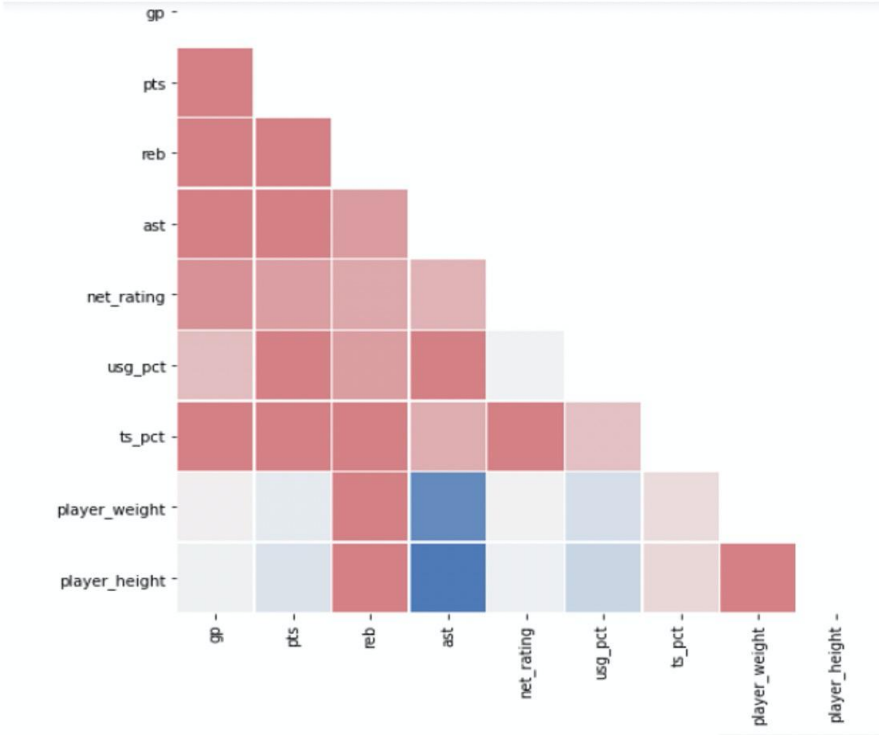
从上面这个宁波2020天气对比饼图可以看出，2020年宁波市主要天气为阴、多云、晴、小雨，其余天气类型占比不是很大。

NBA球员风格变化的分析

身高与体重较高的内线球员在篮板球上明显有充分优势，但他们在助攻与球队命中率上则表现不佳。

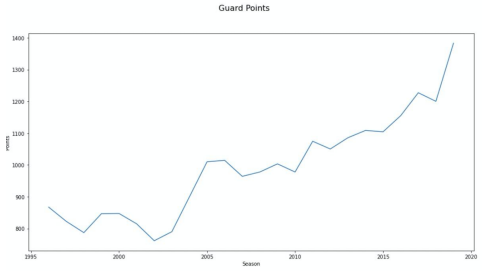
USG_PCT：指的是一个球员在场上时的球队命中率(FGA +控球结束FTA + TO)。

TS_PCT：是衡量球员投篮效率的指标，包括罚球、2分和3分(PTS / (2(FGA + 0.44 FTA)))

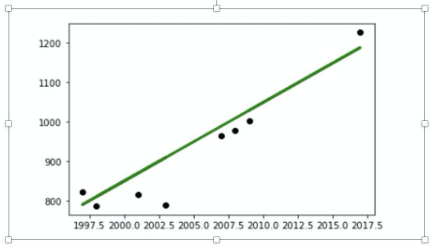


NBA球员风格变化的分析

后卫得分分析
Defender analysis

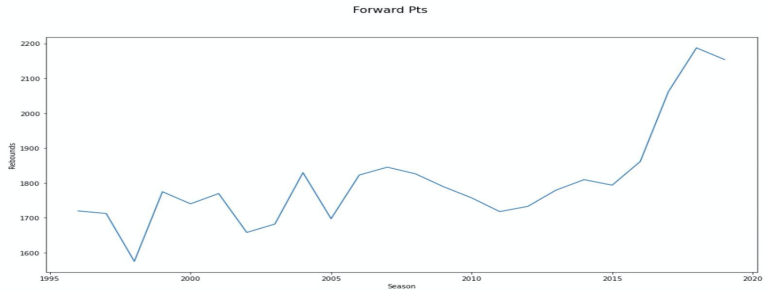


机器学习得分预测



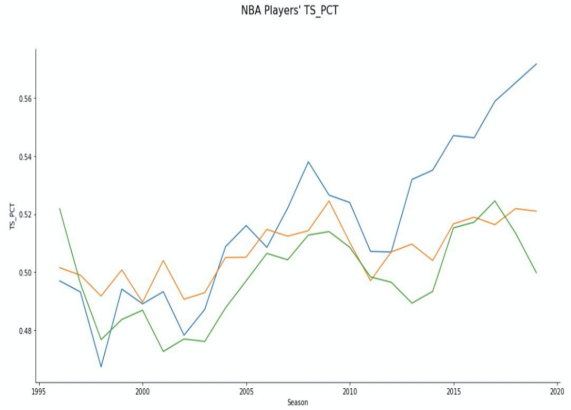
球队更依赖于后卫来得分，这可能是因为他们投三分球，上篮或者扣篮的能力更强。

前锋分析
Forward analysis



这一趋势表明，这些年来，前锋们的得分水平也在不断提高，像勒布朗·詹姆斯和凯文·杜兰特这样的球员就是典型的得分高的前锋。

三位置比对
TS_PCT是衡量球员投篮效率的指标，包括罚球、2分和3分(PTS / (2(FGA + 0.44 FTA)))



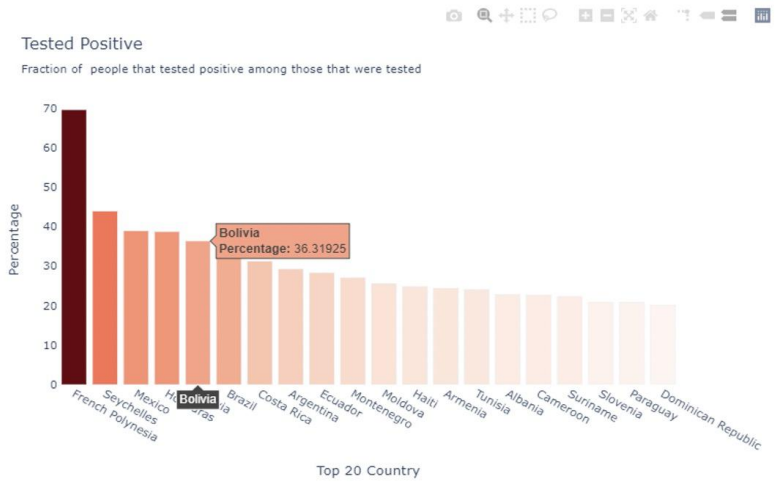
尽管对内线的使用正在减少，但他们的命中率却更高。这可能是由于大多数中锋的得分是灌篮或在篮下投篮的来的高命中率

海豚实验室

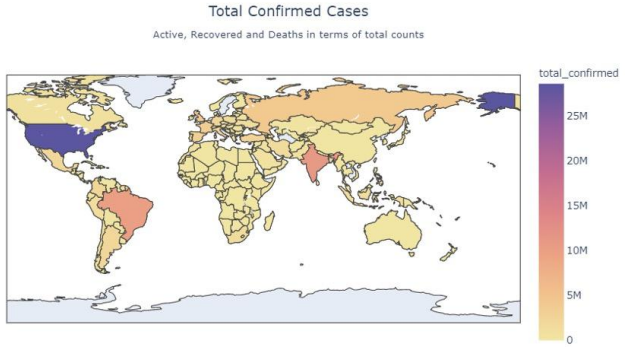
全球疫苗接种率趋势分析

被测人群中被测阳性人群的比例

In [12]:
```
title = get_multi_line_title("Tested Positive ", "Fraction of  people that tested positive among those that were tested")
visualize_column(summary.reset_index(), 'country',"tested_positive", title, "Reds", n=20, ylabel='Percentage' )
```

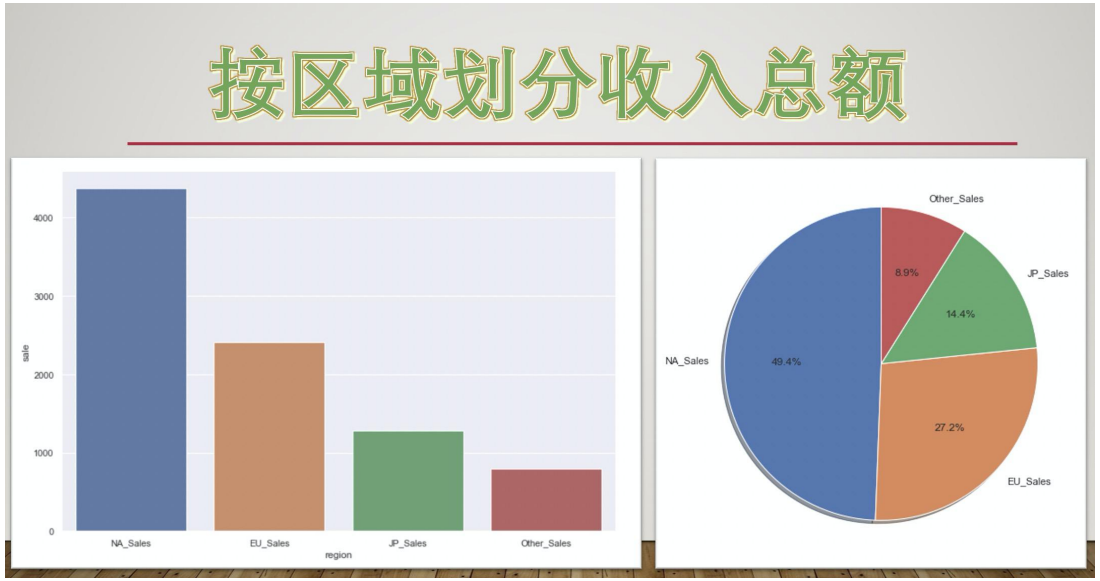


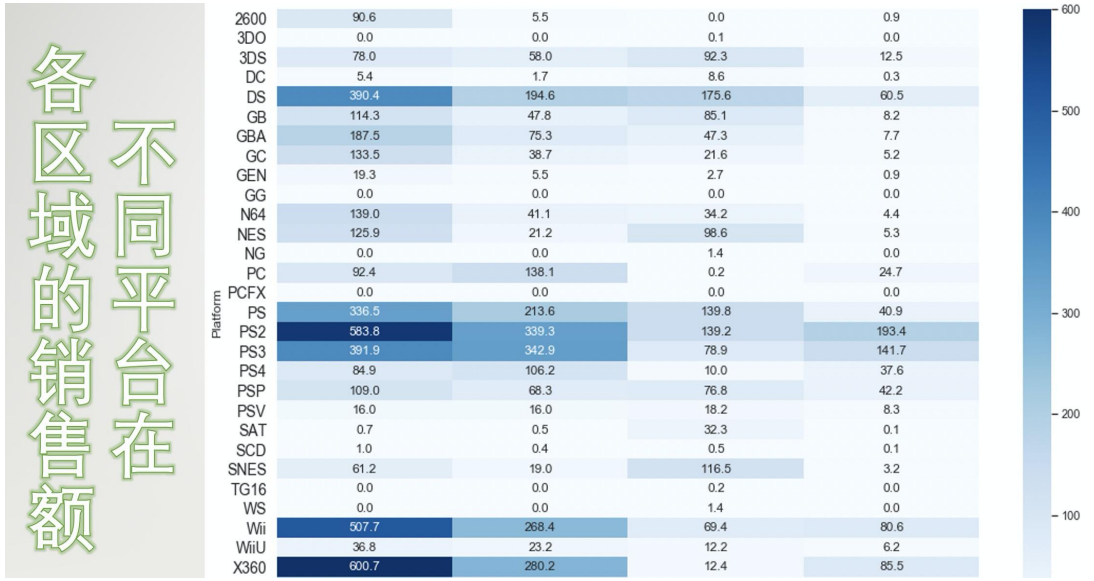
In [17]:
```
# choropleth maps for easier visualization
title = get_multi_line_title("Total Confirmed Cases", "Active, Recovered and Deaths in terms of total counts")

fig = ex.choropleth(data, locations="country",
                    locationmode='country names',
                    color="total_confirmed",
                    hover_name="country",
                    hover_data=['total_deaths', 'active_cases', 'total_recovered'],
                    title='Global Vaccinations over time',
                    color_continuous_scale="Sunset"
                   )

fig.update_layout(title=title,
#                   coloraxis_showscale=False,
                  title_x=0.5)
fig.show()
```



活跃，已恢复和死亡总数

电子游戏销售分析与可视化





新模式：解决课堂中的“三大”痛点

海豚实验室

“兴趣”教学

选题广、数据多、探索式

”互动式“教学

“探讨式”实验案例教学

”能力”教学

项目案例的迭代与重构

专业方向与课程学习的“体感”

典型成功案例（一）

海豚实验室

同济大学 大数据与人工智能实验平台

同济大学

学校情况

同济大学，教育部与国家海洋局、上海市共建的全国重点大学；中央直管副部级建制高校，国家“世界一流大学建设高校”，国家“211工程”和“985工程”建设高校

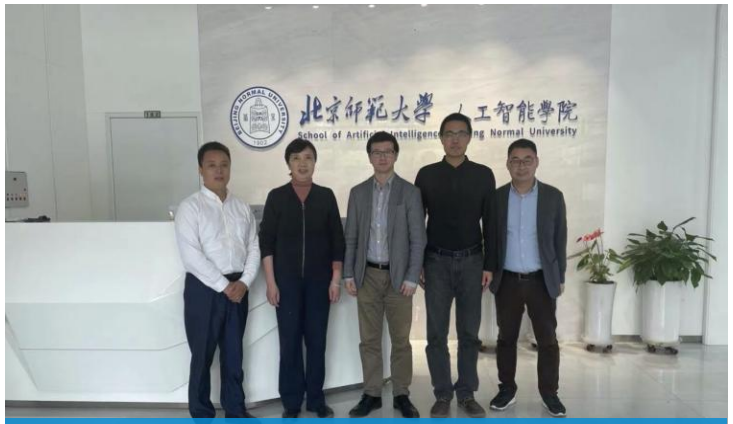
建设内容

为电子与信息工程学院专业基础课、专业核心课程重要的实验支撑平台，服务课程实验、课程设计、项目设计及实训等。

典型成功案例 （二）

海豚实验室

北京师范大学 人工智能实验室



北京师范大学人工智能学院

学校情况

北京师范大学是中华人民共和国教育部直属、教育部与北京市共建的全国重点大学,位列“双一流”、“985工程”、“211工程”。

建设内容

全面支撑北京师范大学人工智能学院提供专业基础课/核心课/选修课的课程教学、实验、课程设计、项目设计、案例实训等。

课程体系、培养范式、项目实战

典型成功案例 （三）

海豚实验室

中国计量大学 大数据分析与人工智能实验室

中国计量大学

学校情况

中国计量大学，是一所以计量、标准、质量、市场监督和检验检疫为办学特色的本科院校。入选教育部“卓越工程教育培养计划”、“新工科研究与实践项目”

建设内容

为信息工程学院提供大数据和人工智能实验室争议解决方案，满足教学、科研的一体化服务，并支撑全校大数据、人工智能相关课程的实验、实训

全面支撑课程实验、实训项目设计、毕业设计等

典型成功案例（四）

海豚实验室

兰州文理学院　大数据实验室

学校情况

兰州文理学院，学校是甘肃省首批转型发展试点院校，文旅部非遗传承人群研修基地，教育部规建中心全国“高等学校产教融合创新实验项目”5个项目院校之一

建设内容

实验室建设包含环境装修、文化墙、服务器、网络、投影、电子白板等各类设备的方案设计与安装部署。共计可支持约53位学生同时上课学习，每周承担近40个学时的课程、实验教学等

全面支撑课程实验、实训项目设计、毕业设计等

海豚实验室“新标准”

PART 04

公司资质

海豚实验室

- “海豚实验室”整体解决方案已获得23项软件著作，2项发明专利申请中
- 入选2021杭州市“雏鹰计划”企业
- 荣获2020全国校企合作十佳案例

- 2020年荣获教育部感谢状
- 2020年全国校企合作十佳案例
- 2019全国高校人工智能与大数据实验室建设领军企业奖
- 2018年获教育部授牌《企业创新训练营》
- 阿里云“人工智能课程”优质内容服务商

理事单位

- 信息技术新工科产学研联盟理事单位
- 中国教育创新校企联盟理事单位
- 中国金融科技教育与应用创新联盟首批理事单位
- 浙江省人工智能学会理事单位
- 全国高校人工智能与大数据创新联盟理事单位
- 阿里巴巴集团合作供应商

公司资质

海豚实验室

教育部产学合作协同育人项目

公司自2019年以来累计获批100多个校企合作项目

2021年7月，杭州睿数科技又获批了共计27个项目建设

企业名称	项目类型	项目简介	项目数量	涉及专业与产业方向
杭州睿数科技有限公司	实践条件和实践基地建设	围绕人工智能、大数据、云计算、金融科技等前沿技术面向计算机科学与技术、软件工程、网络工程、电子商务、信息与计算科学、电子信息工程、计算机应用与维护、数学应用、统计学等相关专业等多个方向推动大学生、教师相关实战能力培养的实践条件和实践基地建设项目。实践教学条件和实践基地建设围绕院校专业教学中的课程实操环节、课程设计环节、实验周环节，以及毕业设计等环节的开展，以专业岗位综合技能培养为目的，以专项解决方案、整体综合解决方案的形式，与院校共同建设相关专业的联合实验室和实践基地。	10	人工智能”、“大数据”、“云计算”、“金融科技”
杭州睿数科技有限公司	创新创业教育改革	面向计算机科学与技术、软件工程、网络工程、电子商务、信息与计算科学、电子信息工程、计算机应用与维护、数学应用、统计学等相关专业等专业建设创新创业教育课程体系，分享创新创业教育最佳实践案例、共建创新创业教育在线课程、创新创业教育实践教材、创新创业教育混合式教学实践、创新创业大赛案例分析与研究、搭建精益创业教育实训基地、举办创新创业教育研讨会等，开展以技术创新为核心的创客教育。	2	计算机科学与技术、软件工程、网络工程、电子商务、信息与计算科学、电子信息工程、计算机应用与维护、数学应用、统计学
杭州睿数科技有限公司	师资培训	面向人工智能、大数据、云计算、金融科技等前沿技术针对高校计算机科学与技术、软件工程、网络工程、电子商务、信息与计算科学、电子信息工程、计算机应用与维护、数学应用、统计学等相关专业教师给予培训名额，采用海豚数据科学一线权威师资、真实项目环境，重点强化实训和项目指导，以提升理论水平、实践能力、创新教学方案为目标，讲授最新技术知识体系并进行技能实训。	15	计算机科学与技术、软件工程、网络工程、电子商务、信息与计算科学、电子信息工程、计算机应用与维护、数学应用、统计学

睿数科技·教育部产学合作协同育人（2021年）项目简介

百校千课免费支持计划

新学期伊始，海豚数据科学实验室特推出”百校千课计划”来支撑全国高校大数据、人工智能等专业的动手实验、课程设计、毕业设计、项目实训等，提供免费云平台和课程、实验资源！

实验课程资源

100+ 课程　800+ 实验　50+ 行业　云服务器资源

你将免费获得：

- 云实验平台教师账号1个+学生账号30个
- 免费云计算服务器资源
- 多门大数据、人工智能课程、实验案例、视频
- 三个行业应用案例，支持课程实验、项目设计、课程设计、毕业设计、实训周等

授课与助教

- 结合课程教学需要，提供18学时直播授课
- 结合实验实训要求，提供36学时助教辅导

扫码申请

大数据/人工智能专建设：三大痛点

海豚实验室

1

新技术进入课堂

课堂教学不同于技术培训
教师需要积累和沉淀，创建合适
的大纲和教学内容

2

缺乏教学案例

学生对行业认知有限，兴趣不足，
职业规划不清
极度缺乏真实的应用案例和数据集

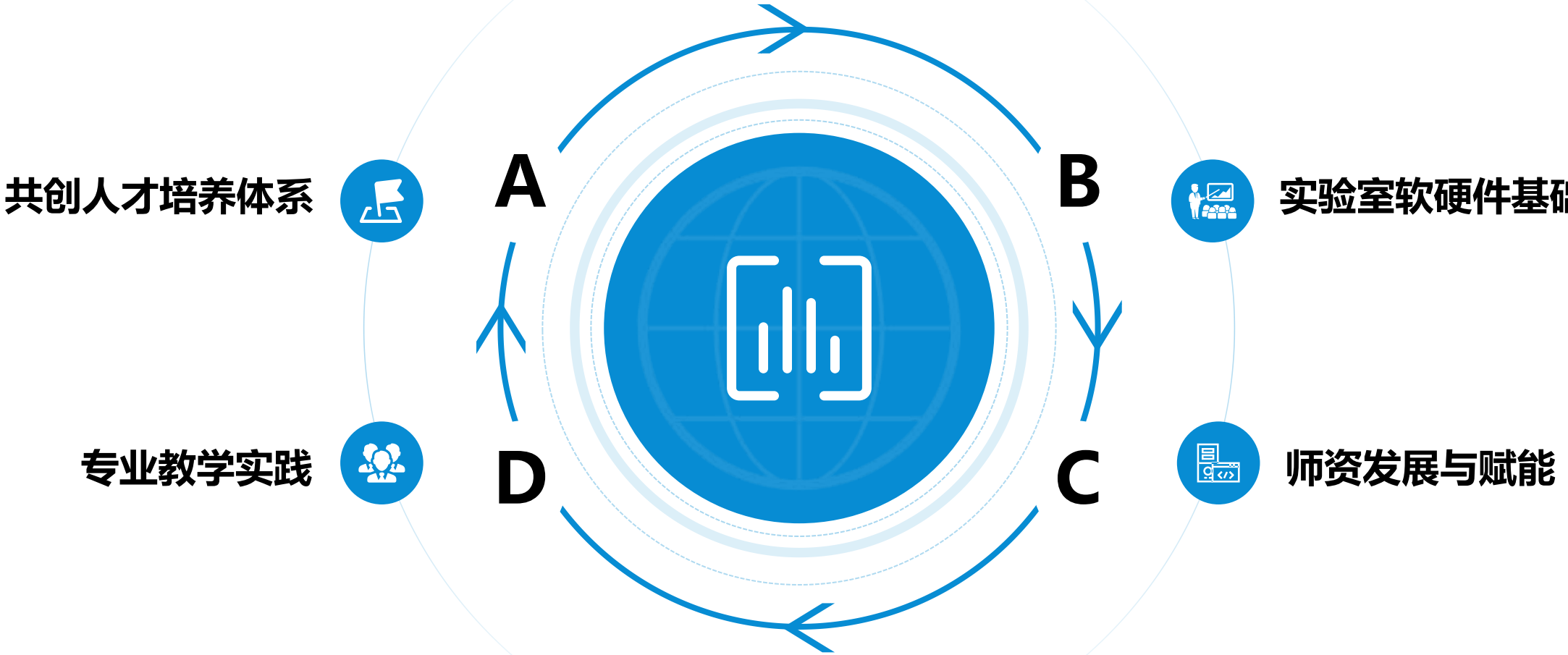
3

缺乏一体化平台

系统复杂，且资源要求高
教学与科研系统相对分离，良性
循环难以实现

海豚实验室：“可持续性赋能”的新标准

海豚实验室



平台 + 内容 + 服务

海豚教育解决方案

海豚实验室

实验实训平台

- 虚拟桌面
- 班级管理
- 案例库资源

竞赛平台

- Juypter环境
- 课程管理
- 数据集资源

工程认证平台

- RStudio工具
- 考评管理
- 智能分析

科研平台

- 教学视频
- 竞赛题库
- 容器管理
- 教学课件
- 项目实训
- 集群资源管理

产品特色

海豚实验室

350+ 课程

3000+ 实验

350+ 行业案例

5TB+ 数据集

1

“口袋里”的实验室

- 基于B/S全浏览器模式
- 无边界实验室，边学边练
- 独占实验资源，智能调度

2

精品课程/实验资源

- 专业培养大纲无缝对接
- 真实行业案例、数据集资源
- 提供定制化课程/实验开发

3

企业创新训练营

- 教育部领导授牌认可
- 学生/师资训练营服务
- PDL项目式教学实践

大数据/人工智能课程教学+实验实训+科研创新

海豚实验室

实验实训资源

预配实验时序环境	算法模型	数据集
PyTorch TensorFlow Keras Pandas Sclklt-learn Bommlearn OpenCV NLTK Jieba Gensim Librosa ……	聚类 回归 决策树 支持向量机 VGG CNN RNN GAN FCN HMM ……	航班客流 新闻关键词 手写数字识别 物体分类检测 人脸识别 表情年龄属性 …… **3000+ 实验** **(支持Jupyter Notebook格式)**

理实一体

专业课程体系

(标*的课程资源提供TansorFlow与pytorch两种版本)

神经网络与深度学校*	OpenCV图像处理	知识图谱
机器学习	深度学习与计算机视觉*	语音信号处理
Python数据分析	TensorFlow深度学习实战*	自然语言处理*
1500+课件PPT **3000+习题** **教学大纲**	人工智能概论	人工智能数学基础

AI实验实训

AI教学支持

教学实训平台

- 科研底层支持
- 实验实训环境
- 教学过程跟踪
- 课程资源管理
- GPU资源调度

专业建设服务：师资研修 | 训练营 | 实习服务

服务支撑

端云结合场景式AI应用

AI应用开发设备

- 场景式人工智能应用开发箱
- ROS无人车
- 服务机器人

人脸识别 表情年龄属性 物体分类检测 ……	语音识别 语音合成 智能对话 ……	SLAM 自动导航 自动跟随 智能驾驶 ……

AI算力支持

AI算力支持

专业基础硬件设施：GPU训练服务器 | 实验实训支撑服务器 | 教学管理服务器

“一站式”大数据/人工智能专业课程资源+实验实训资源+AI设备+科研创新支持
满足高校专业建设需求，培养从事与大数据/人工智能相关的设计型、工程型、应用型人才。

特色1：来自行业案例全面、丰富，适用广！

海豚实验室

学习内容

视频 + 教材 + 动手实验

基础能力+专业工程能力

学习模式

边学边做， 事半功倍

适用于

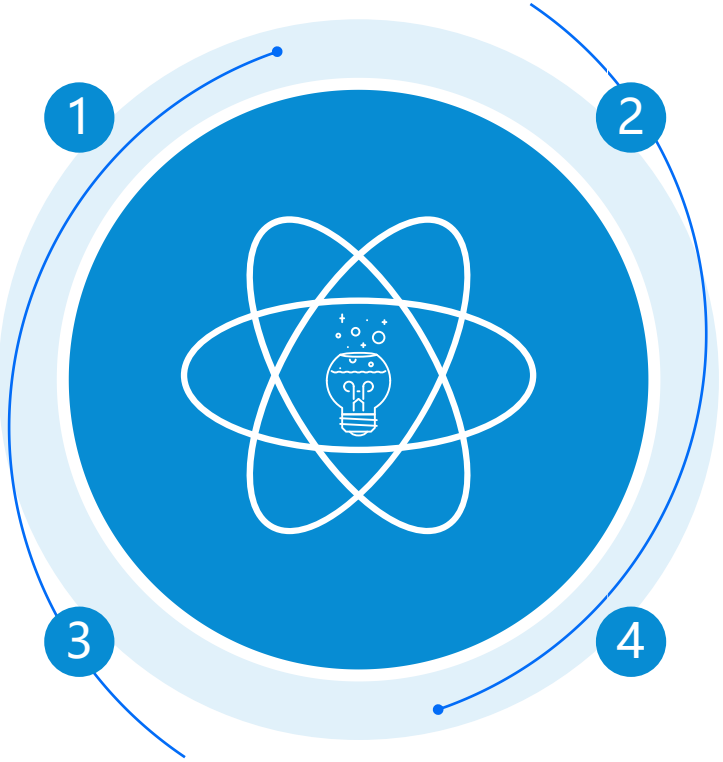
- 计算机
- 数学与统计
- 信息工程
  智能制造
- 金融科技
- 经管类
- ……

特色2：“口袋里”的实验室

海豚实验室

1. 全浏览器访问，实验室/教室/宿舍，随时随地学习

2. 采用智能容器集群+私有云底座，节省70%以上的硬件资源



3. 每个学生独立环境，进度自由保存

4. “教 - 学 - 练 - 评 - 赛 - 研”一体化

特色3：开放式平台与定制化服务

海豚实验室

跨平台、多语言、多环境支撑

(Jupyter，虚拟机，IDE、 CANVAS工具等)

(Java/Javascript，C/C++, SQL,Python,R ,SQL,XSQL,SAS等十几种语言)

支持完全“自定义”的课程、实验等二次开发

提供定制化课程、实验、实训、项目设计等内容开发

二十多种开发工具与实验集成环境

Hadoop 集群
分布式资源计算平台

Hadoop + Spark 集群
高性能分布式资源计算平台

Hadoop + 人工智能 集群
人工智能分布式资源计算平台

Hadoop + 数据存储集群
分布式数据库管理应用平台

Jupyter
对数据建模、训练等分析的数据科学工具

Theia
云IDE多语言编辑器(Theia)

Jupyter-Torch
Torch1.7环境，对数据建模、训练等分析的数据科学工具

Jupyter-Tensorflow
Tensorflow2.0环境，对数据建模、训练等分析的数据科学工具

RStudio
对数据进行统计、分析的工具

Seahorse
支持拖拉拽的可视化数据分析工具

OpenRefine
对数据进行预处理、提炼的工具

Excel
表格数据处理软件

PowerBI
创建和共享交互式数据可视化

SAS
主流的商业数据分析工具

Tableau
可视化商业智能数据分析工具

全开放式课程实验体系的自建与定制开发

实　验

- 提供模板创建（1000+个实验模板）
- 提供集成实验工具和环境（Jupyter/虚拟机/IDE等20+个工具与语言）
- 提供数据集仓库导入
- 完美底层封装，专注内容开发

课　程

- 预置实验库 + 自定义实验
- 预置课件库 + 自定义课件
- 预置视频库 + 自定义视频
- 预置习题库 + 自建习题

课程体系

- 预置课程+自建课程
- 学习路径
- 行业岗位匹配（阿里、百度、腾讯、工信部等证书）

特色4：岗位能力画像、产教融合大屏

海豚实验室

“反向式”的课程、
实验内容设计

岗位
能力匹配

产教融合
大屏

产教融合实时大屏

海豚实验室

学生中心

海豚实验室

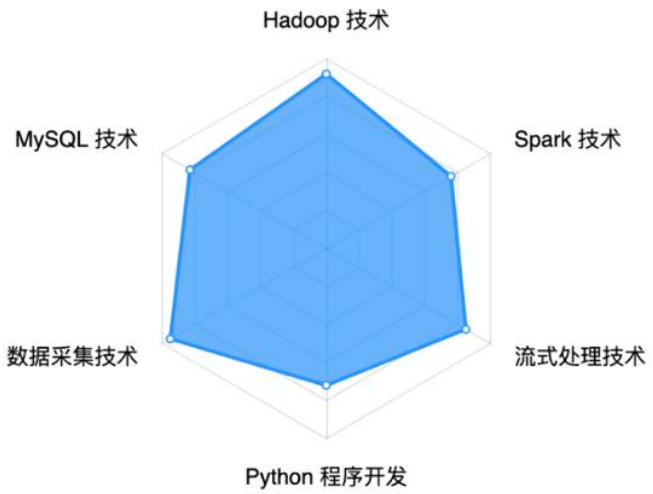
support
2017-11-15 加入海豚实验室
累计实验次数 3237 次
有效学习时长 97110 分钟
学习概览
我的班级
我的课程
我的数据
我的考试
我的作业
我的岗位
邀请好友
发送下面链接给好友
https://labs.wd.dolphin-labs.con
复制
分享至
学习中心 / 学习概览
中
良
差
优
综合分数
596
报告得分 | 20% 109
考试得分 | 50% 348
作业得分 | 20% 87
平时得分 | 10% 52
报告得分
报告完成度+得分 109
70%
当前排名:7/16
考试得分
考试总成绩 348
65%
当前排名:9/16
作业得分
作业总成绩 87
30%
当前排名:3/60
平时表现得分
平时表现得分 52
20%
当前排名:10/60

学生中心2

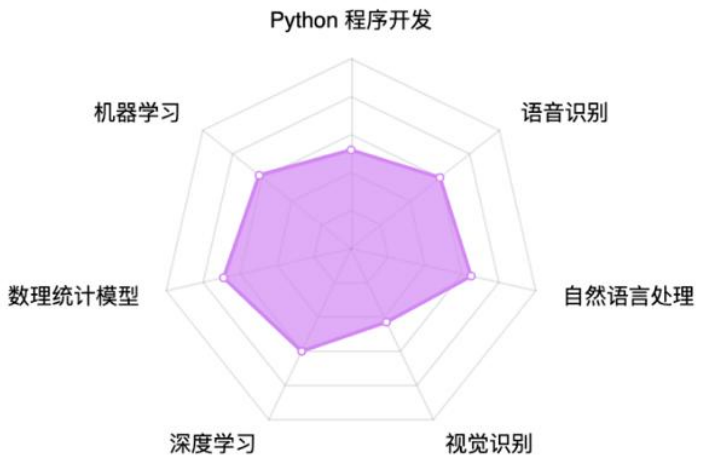
海豚实验室

岗位能力匹配

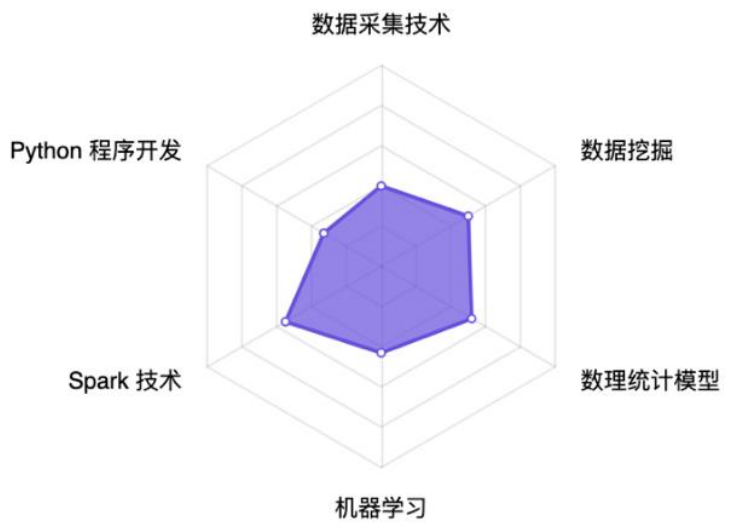
数据工程师 匹配值:80%



数据科学家 匹配值:60%

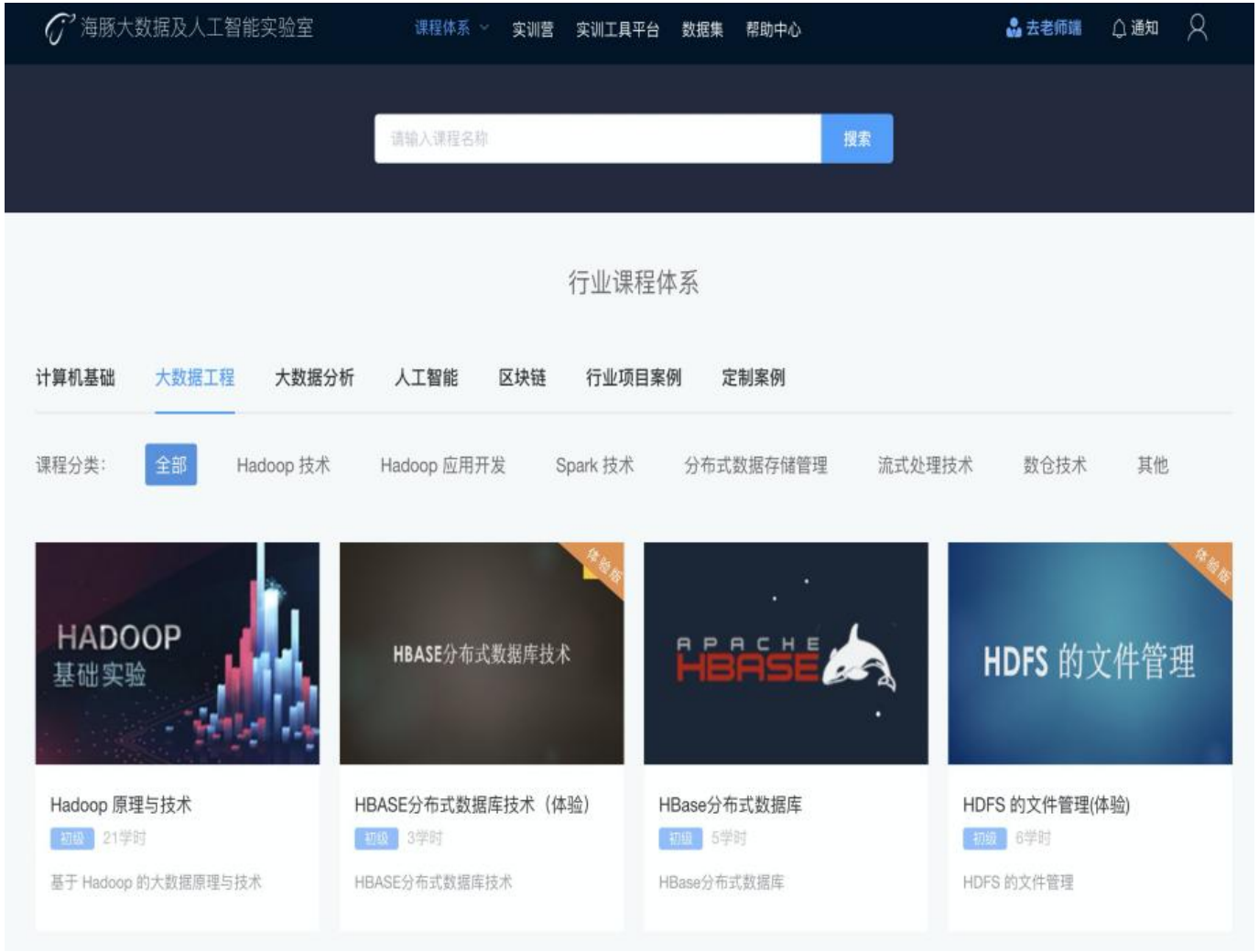


数据分析师 匹配值:30%



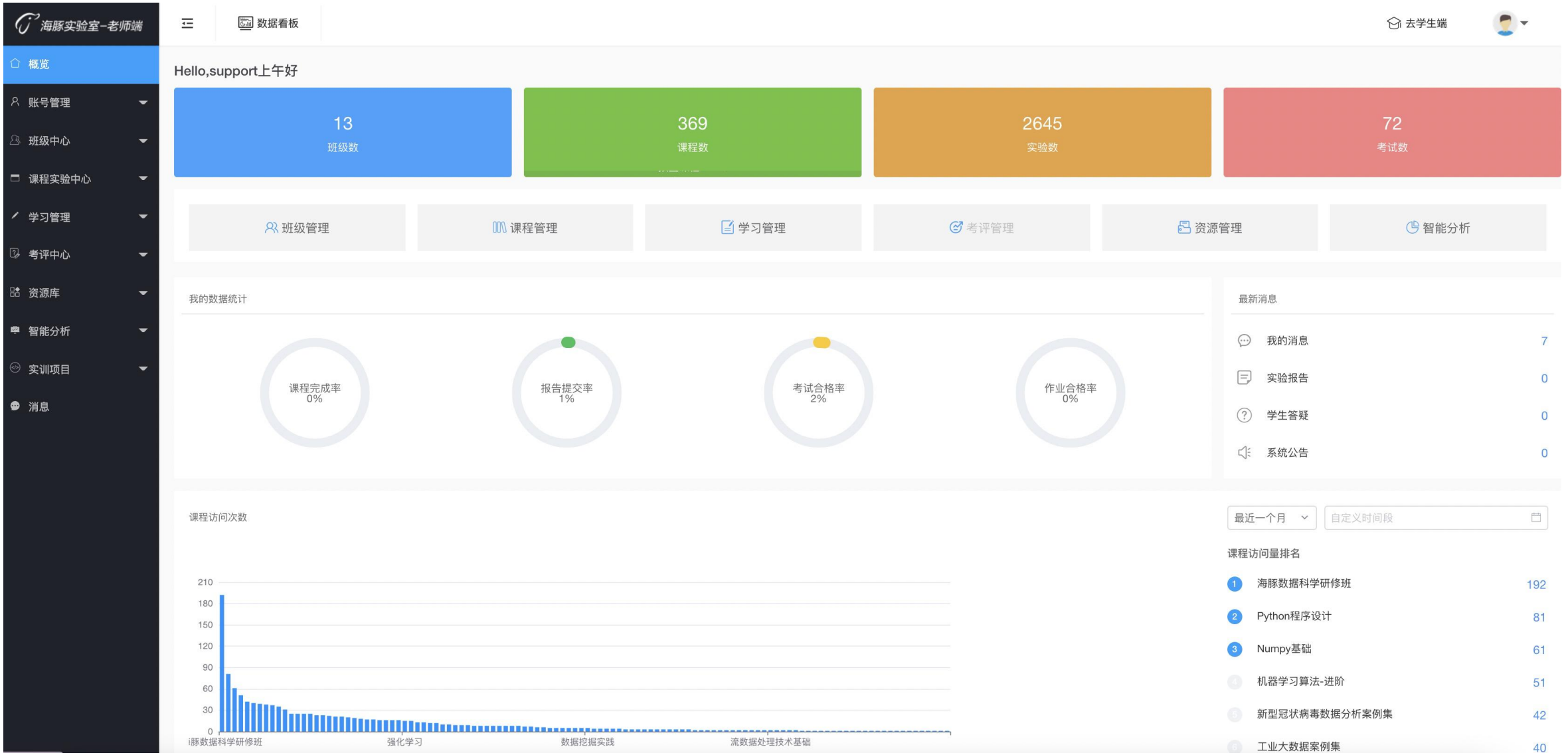
实验实训平台

海豚实验室



教学及管理平台

海豚实验室



赛事规划

海豚实验室

“以赛促学、以赛促教”

校赛

校内赛

参与人员

面向本院系及本校各相关专业学生举办

省赛

区域高校赛

参与人员

面向本省或本取悦高校相关专业学生

国赛

全国院校赛

参与人员

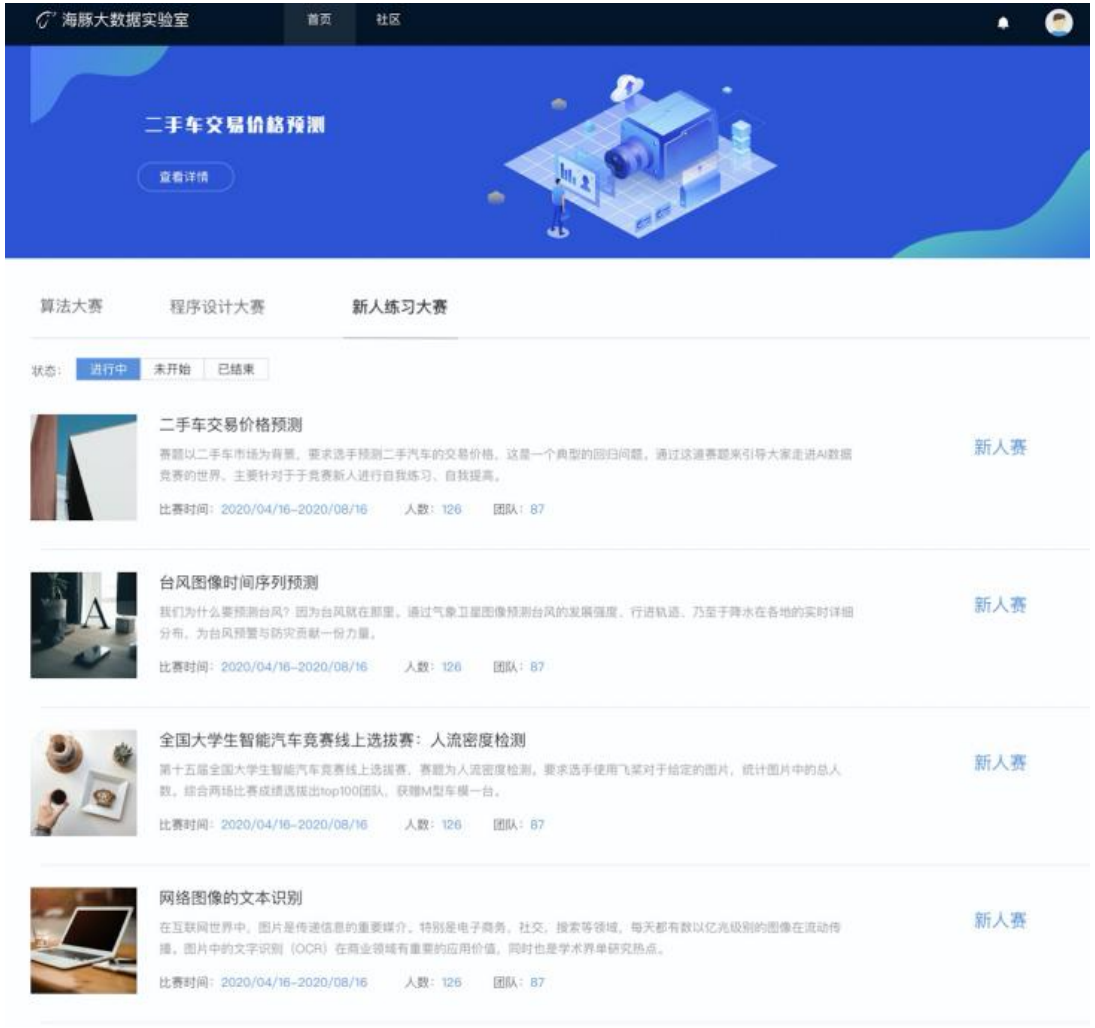
面向全国各大院校相关专业大学生、科研单位人员等

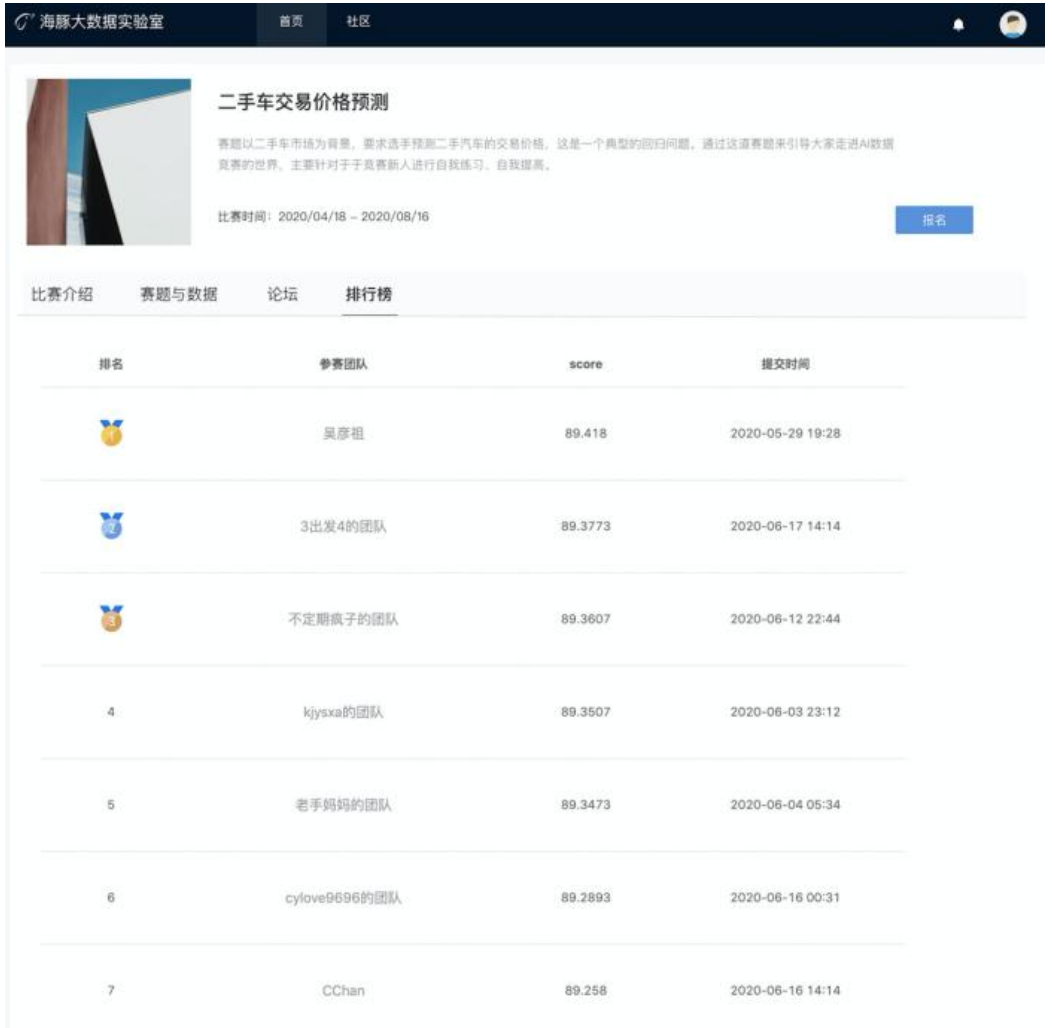
赛事内容： 知识赛 大数据分析赛 大数据应用创新赛

全面提升学校地区影响力、服务能力，为培养具备大数据、人工智能主流技术的新一代信息技术人才。

竞赛平台

海豚实验室


海豚大数据实验室
首页
社区
二手车交易价格预测
查看详情
算法大赛
程序设计大赛
新人练习大赛
状态：
进行中
未开始
已结束
二手车交易价格预测
赛题以二手车市场为背景，要求选手预测二手汽车的交易价格，这是一个典型的回归问题，通过这道赛题来引导大家走进AI数据竞赛的世界，主要针对于于竞赛新人进行自我练习、自我提高。
比赛时间：2020/04/16-2020/08/16
人数：126
团队：87
新人赛
台风图像时间序列预测
我们为什么要预测台风？因为台风就在那里，通过气象卫星图像预测台风的发展强度、行进轨迹、乃至于降水在各地的实时详细分布，为台风预警与防灾贡献一份力量。
比赛时间：2020/04/16-2020/08/16
人数：126
团队：87
新人赛
全国大学生智能汽车竞赛线上选拔赛：人流密度检测
第十五届全国大学生智能汽车竞赛线上选拔赛，赛题为人流密度检测，要求选手使用飞桨对于给定的图片，统计图片中的总人数，综合两场比赛成绩选拔出top100团队，获赠M型车模一台。
比赛时间：2020/04/16-2020/08/16
人数：126
团队：87
新人赛
网络图像的文本识别
在互联网世界中，图片是传递信息的重要媒介，特别是电子商务、社交、搜索等领域，每天都有数以亿兆级别的图像在流动传播，图片中的文字识别（OCR）在商业领域有重要的应用价值，同时也是学术界单研究热点。
比赛时间：2020/04/16-2020/08/16
人数：126
团队：87
新人赛


海豚大数据实验室
首页
社区
二手车交易价格预测
赛题以二手车市场为背景，要求选手预测二手汽车的交易价格，这是一个典型的回归问题，通过这道赛题来引导大家走进AI数据竞赛的世界，主要针对于于竞赛新人进行自我练习、自我提高。
比赛时间：2020/04/18 - 2020/08/16
报名
比赛介绍
赛题与数据
论坛
排行榜
排名
参赛团队
score
提交时间
1
吴彦祖
89.418
2020-05-29 19:28
2
3出发4的团队
89.3773
2020-06-17 14:14
3
不定期疯子的团队
89.3607
2020-06-12 22:44
4
kjysxa的团队
89.3507
2020-06-03 23:12
5
老手妈妈的团队
89.3473
2020-06-04 05:34
6
cylove9696的团队
89.2893
2020-06-16 00:31
7
CChan
89.258
2020-06-16 14:14

工程教育认证平台

海豚实验室

大数据及人工智能行业案例覆盖

海豚实验室

- 工业制造与大数据分析方案
- 农业大数据应用与实践
- 医疗大数据案例与分析
- 零售大数据产业分析与案例分享
- 大数据与工业互联网
- 财税大数据案例与分析
- 政务大数据案例与分析
- 运营商大数据案例与分析
- 餐饮行业大数据分析方案
- 旅游大数据体系案例与分析
- 招商大数据案例与分析
- 教育大数据解决方案
- 海关大数据案例与分析
- 金融大数据产业与案例
- 扶贫大数据建设方案
- 电商零售大数据分析方案
- 化工行业大数据分析方案
- 地产大数据分析方案

金融科技案例集 (体验)

高级 9学时

以金融行业应用的场景和需求为背景，大数据及人工智能技术应用的实验案例集 ...

工业大数据案例集(体验)

初级 2学时

工业大数据案例集(体验版)

数据科学应用演示集(体验)

初级 8学时

本课程致力于体验设计交互式NoteBook课程模式，以及探索大数据分析以及人 ...

电信行业案例集（体验）

初级 6学时

电信行业的大数据分析应用案例集

新闻媒体类案例集(体验)

初级 6学时

新闻媒体类案例集

经济管理类案例集(体验)

初级 6学时

经济管理类案例集

健康与医疗类案例集(体验)

初级 7学时

健康与医疗类案例集

军事大数据案例集(体验)

初级 2学时

军事大数据的分析案例

数据集

海豚实验室

提供行业数据集包不少于30个行业，包括金融，电信，互联网、房地产、电子商务、智能制造、旅游出行、汽车、保险、交通等行业数据集，数据总容量不低于**5TB**。部分内容如下：

房产数据 01

汽车数据 02

股票基金交易数据 03

环境与气候数据 04

人脸图像数据 05

行为视频数据 06

搜索引擎访问日志数据 07

海量网站访问日志数据 08

政务数据 09

电商交易数据 10

工业制造数据 11

旅游数据 12

海豚实验室

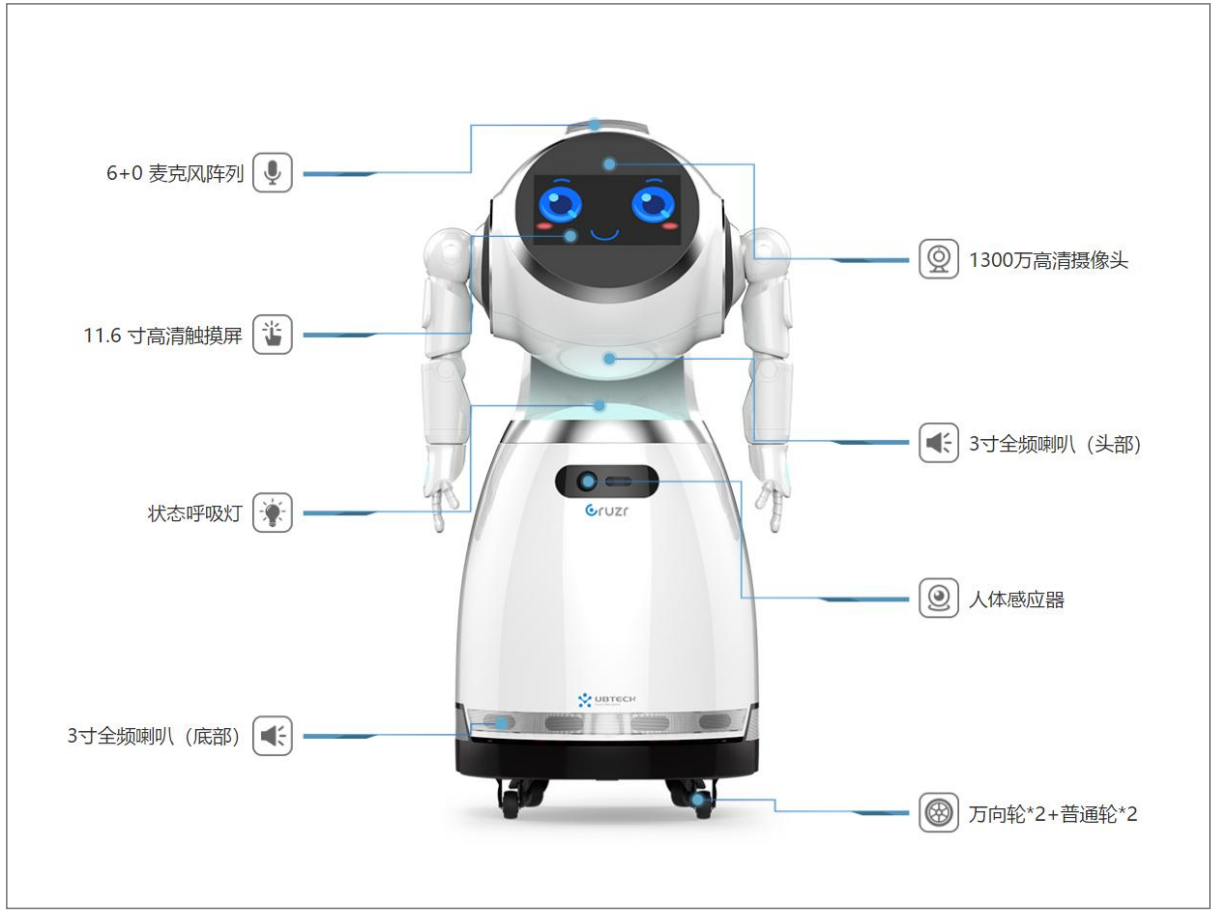
方案一：人工智能套件

- 语音识别控制机械臂
- 视觉分拣工作站
- 人工智能语音套件
- 基于ROS的移动协作机械臂开发平台
- 人工智能技术应用基础实验台

海豚实验室

方案二：智能服务机器人

- 多模态人机交互体验
- 精准人脸识别
- U-Meeting视频客服系统
- 多台设备一键多控
- 主动式数据采集管理
- 电子皮肤安全保护



产品及服务

海豚实验室

1

云服务

课程资源
云账号

2

实验室建设

大数据实验室
人工智能实验室

3

培训与教育服务

师资研修班
学生训练营

解决方案

海豚实验室

工程认证

课程大厅	课程管理	实验手册	题库	课题报告	课程审核	项目协作	智能资源调度
Hadoop编程	MySQL数据库	Spark技术	深度学习	自然语言处理	语音识别		
Python语言	机器学习	数据可视化	金融科技案例	电子商务案例	智能制造案例		

竞赛

实验实训	教学管理	二次开发引擎	数据采集	数据标注	Spark引擎	数据驾驶舱
作业考评	报告管理	自定义开发	模型训练	视觉识别引擎	自然语言引擎	Hadoop集群

实验实训

科研支撑与管理

数据管理

交通数据
金融数据
电商数据
政务数据
……

计算资源底座

智能容器集群			智能虚拟机集群		
存储资源池	CPU计算资源池	GPU计算资源池	网络资源池	VLAN池	……

云服务在线课程

海豚实验室

模块名称	大数据工程	大数据分析	人工智能	行业案例	账号数目
标准版	任选 2 门 实验课程	任选 2 门 实验课程	N/A	任选 2 个 行业案例	<= 40 个
进阶版	任选 3 门 实验课程	任选 3 门 实验课程	任选 3 门 实验课程	任选 3个 行业案例	<= 40 个
高级版	任选 5 门 实验课程	任选 5 门 实验课程	任选 5 门 实验课程	任选 5 个 行业案例	<= 40 个

实验室解决方案 – 本地部署

海豚实验室

模块名称	大数据工程	大数据分析	人工智能	科研及资源管理	并发数	授权账号数目	维保
高职版	✓	✓			<= 100 人	无上限(需硬件支持)	2 年
本科版	✓	✓	✓		<= 100 人	无上限(需硬件支持)	3 年
科研版	✓	✓	✓	✓	<= 100 人	无上限(需硬件支持)	4 年

学生训练营

海豚实验室

模块名称	学时	学生数量	师资服务
标准版	36学时 (6天)	<=50个	1名讲师
进阶版	48学时 (8天)	<=50个	1名讲师
高级版	60学时 (10天)	<=50个	1名讲师+1名助教

大数据实验室 - 大数据工程

海豚实验室

大数据工程基础实验	大数据工程中级实验	大数据工程行业案例	区块链技术
Hadoop单点安装和部署	数据迁移到Hadoop	**基于某电商平台访问日志的数据分析**	以太坊的安装和部署
Hadoop集群安装和部署	用Zookeerper开发分布式应用	**基于微信朋友圈的用户画像分析**	以太坊智能合约开发实践
Ambari部署和管理	Spark集群的安装和部署	**基于某城市出租车GPS位置数据的分析**	Fabric1.0集群的安装和部署
Pig的部署和管理	Pyspark的应用开发	电商网站商品推荐的数据分析和处理	创建你自己的加密货币实验
Zookper的部署和管理	SpakR的应用开发	基于世界银行公开数据的分析	
MapReduce的编程基础	Spak stream的应用开发	白血病医学文献的文本分析	
Spark的单点安装和部署	Storm的安装和部署		
Cassandra的安装和部署	基于Storm的流处理分析		
MongoDB的安装和部署	Apache Lucene的安装和部署		
	Apache Solr的安装和部署		
	PyLucene的安装和部署		

大数据实验室 - 数据分析

海豚实验室

大数据分析基础实验	大数据分析进阶实验	大数据分析行业案例	机器学习算法实验
Python编程基础	Python语言详解	基于Python的爬虫应用实践	基于Python的机器学习基础
R编程基础	R语言详解	股票标准普尔500走势的数据分析	基于R的机器学习基础
基于Python的数据分析	Scala语言详解	基于中国气象站的气候数据分析	K-means 算法
基于Python的数据统计	Pandas的详解	泰坦尼克号受难人员的数据分析	逻辑回归算法的基础
基于Python的数据可视化	**Numpy的数据分析详解**	股票指数走势与社交媒体关联度分析	分类/聚类算法
基于R的数据可视化	Matplotlib 可视化详解	基于OpenCV的人脸识别	决策树算法
	Scikit-learn 数据分析详解	航班晚点的预测分析	支持向量机算法
		语音的识别与分析	随机森林算法
		2012年美国大选的民调预测数据分析	协同过滤算法

人工智能实验室-基础实验

海豚实验室

人工智能基础实验	人工智能进阶实验	视觉识别	基于GPU的神经网络
Tensorflow 的基础入门	深度学习Keras框架基础	基于整幅人脸图像的识别算法	生物神经网络
Keras框架应用	贝叶斯分类算法	基于神经网络的识别算法	反馈神经网络
Caffe 的基础入门	递归神经网络	基于支持向量机的识别算法	单层神经网络
PyTorch 基础入门	使用RNN/LSTM生成文本	隐马尔科夫模型	DL多层神经网络
OpenCV基础入门	限制玻尔兹曼机	支持向量机	DBN深度信念网络
CNN卷积神经网络	深度信念网络	人工神经网络	CNN卷积神经网络
RNN循环神经网络		小波变量算法	
NLP自然语言处理		基因编码	

全国高校师资研修班系列

海豚实验室

杭州睿数科技CEO 吴仲毓

www.dolphin-labs.com

400-001-3538
sales@dolphin-labs.com