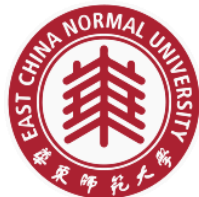


中国科学院学部-科学与技术前沿论坛
大模型 / AIGC 的健康发展与赋能赋智



大模型赋能全民数字素养与技能提升

王伟

国家级全民数字素养与技能培训中心
华东师范大学 / 数据科学与工程学院

驱动力 | 智能时代 |



ALL Systems Go

At last — a computer program that can beat a champion Go player

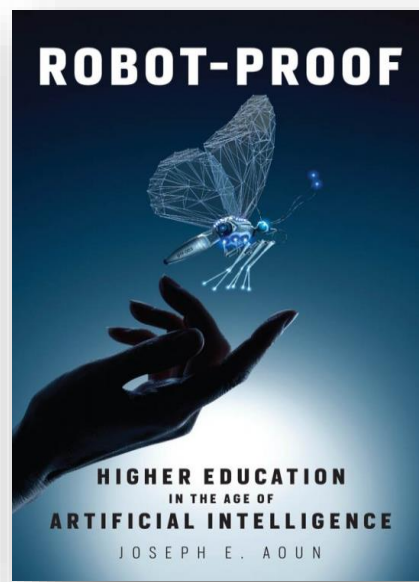
Nature 2016.01



DARK FACTORY

The robotics revolution is changing what machines can do. Where do humans fit in?

The New Yorker 2017.10



Robot-Proof

Higher Education in the Age of Artificial Intelligence

MIT Press 2017.08



AlphaFold

AlphaFold, software that can predict the 3D shape of proteins, is already changing biology.

Nature 2021.08



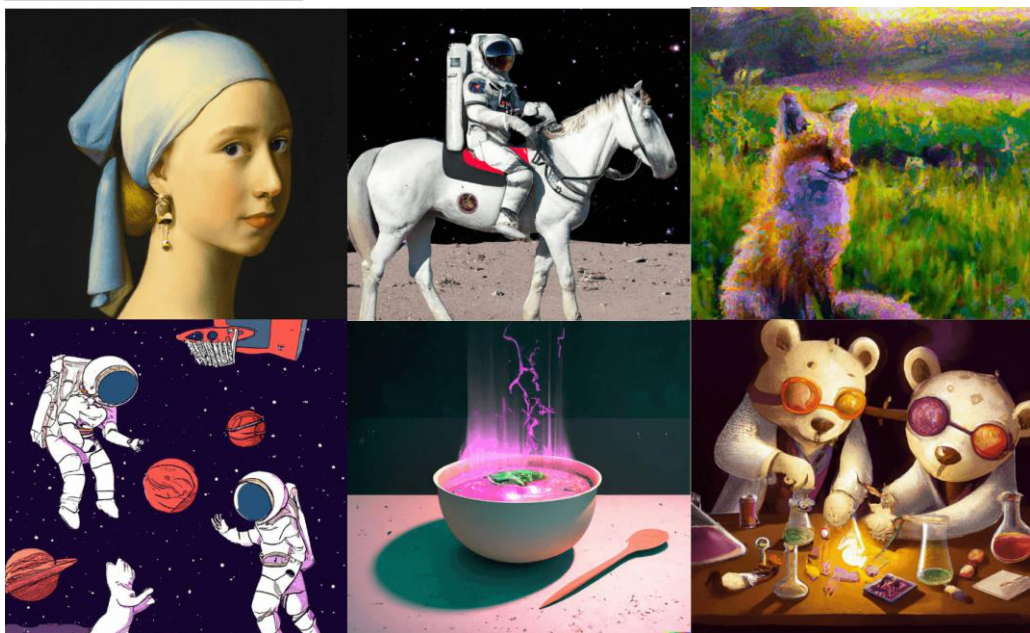
ChatGPT

Ten people (and one non-human) who helped shape science in 2023.

Nature 2023.12

驱动力 | 智能时代 II

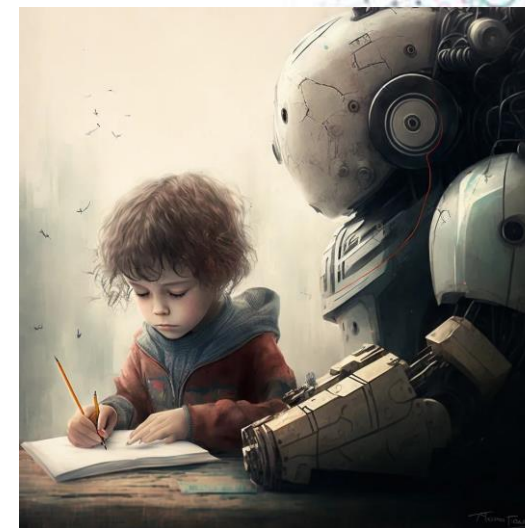
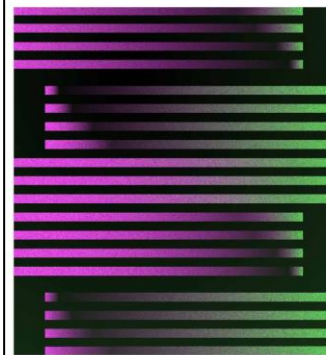
DALL·E



Noah Smith captured the human-AI dynamic succinctly in his “sandwich” workflow:

“This is a three-step process. First, a human has a creative impulse, and gives the AI a prompt. The AI then generates a menu of options. The human then chooses an option, edits it, and adds any touches they like.”

ChatGPT: Optimizing Language Models for Dialogue



 [vitalik.eth](#) 
[@VitalikButerin](#) · [Follow](#)



I can easily see many jobs in the next 10-20 years changing their workflow to "human describes, AI builds, human debugs".

智能时代的我们：**人类描述** → **AI 构建** → **人类调试**
不被 AI 取代的必须：AI 不知道的事情（**专有**），理解 AI 不能理解的东西（**解释性**），制作 AI 还不知道的东西（**创造性**）。

航天学的启示

- ◆ 是**航空技术**而不是“人工飞行”
- ◆ 空气动力学、驾驶舱与仪表盘
- ◆ 技术智能也不应该是“人工的”，
应该是**增强人类能力的智能**



智能时代下的教育

人类增强智能 = 人脑智能 + 技术智能

- 基本素养的提升

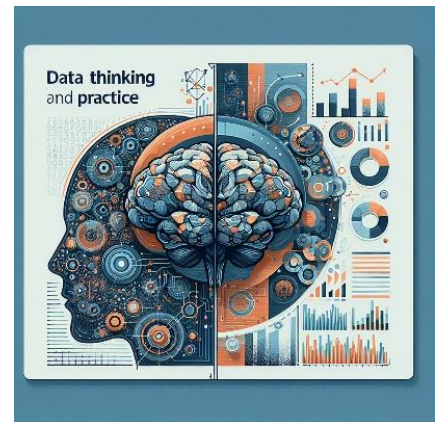
- 数字素养 (Digital literacy) 、 数据素养 (Data literacy) 、 人文素养 (Human literacy)

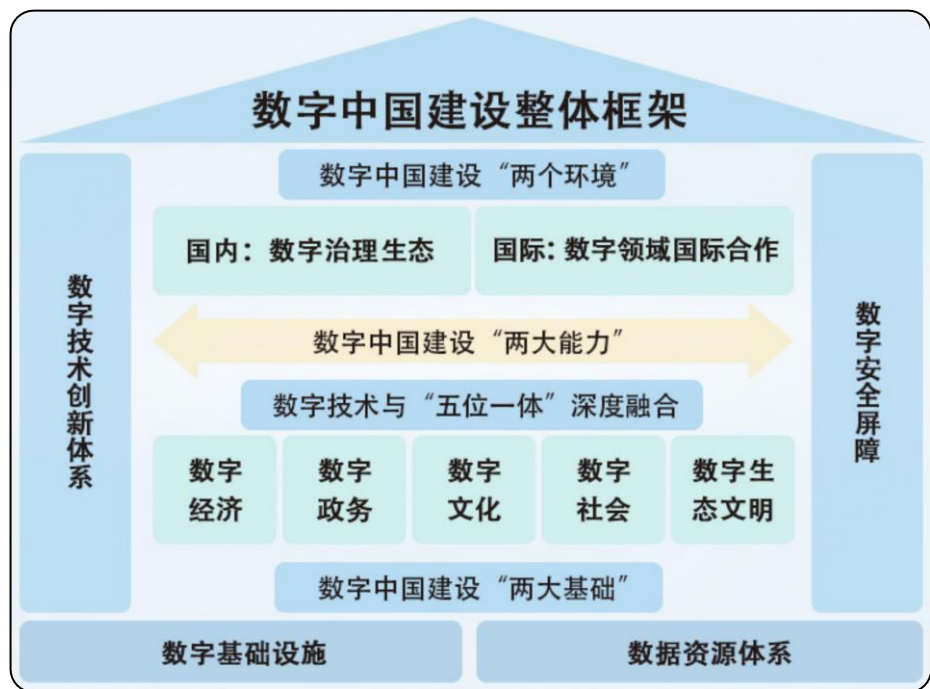
- 核心能力的提升

- 学习能力、问题求解能力、信息获取能力、分析推理能力、决策能力 等等

- 综合认知的提升

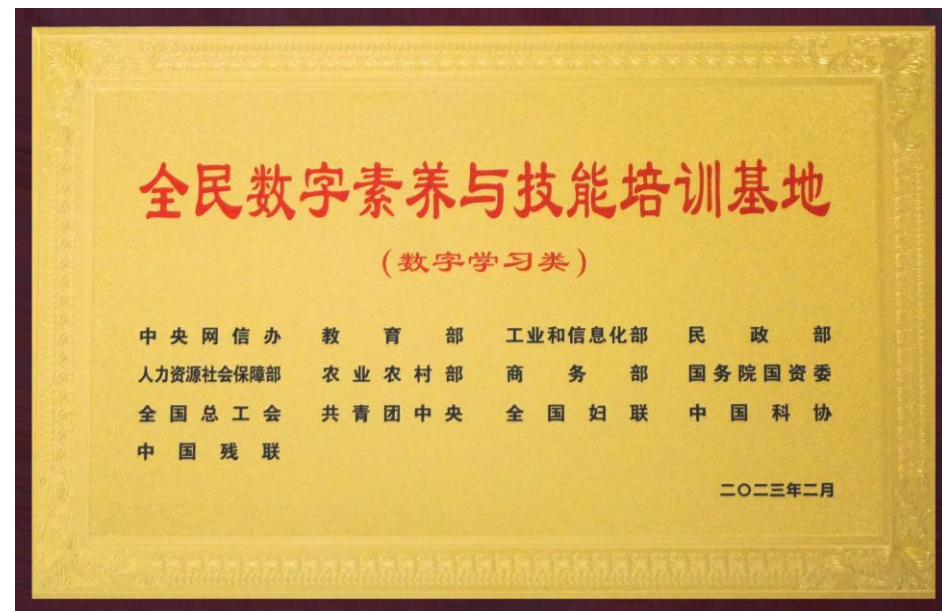
- 系统性思维 (System thinking) 、 数据思维 (Data thinking)
- 设计思维 (Design thinking) 、 批判性思维 (Critical thinking)
- 认知敏捷性 (Cognitive agility) 、 创业精神 (Entrepreneurship)





写入“十四五”规划和
二〇三五远景目标的重要蓝图





华东师范大学 · 全民数字素养与技能培训基地

引进优质社会人才与资源

面向全民提供数字素养与技能提升教育服务

数字素养框架与标准

(2021)

0. 设备和软件操作

1. 信息和数据素养

2. 沟通与协作

3. 创造数字内容

4. 安全

5. 问题解决

6. 职业相关的素养



欧洲
委员会

DigComp
Framework

(2013)



联合国
教科文
组织

Digital
Literacy
Global
Framework

(2018)



(2023.5)

数字素养框架

华东师范大学·全民数字素养与技能培训基地

<2023-05-24 Wed>

目录

1 引言	2
1.1 数字时代的教育挑战	2
1.2 数字化赋能的终身学习	3
1.3 本文结构与用法	3
2 领域	4
2.1 [0] 通用数字设备与应用软件	4
2.1.1 [0.1] 使用智能电子设备	4
2.1.2 [0.2] 使用通用计算机设备	4
2.1.3 [0.3] 使用常用应用软件	5
2.2 [1] 信息与数据	5
2.2.1 [1.1] 浏览、搜索和筛选信息与数据	5
2.2.2 [1.2] 分析、比较和评价信息与数据	5
2.2.3 [1.3] 管理信息与数据	5
2.3 [2] 沟通与协作	5
2.3.1 [2.1] 管理数字身份	5
2.3.2 [2.2] 使用数字技术互动	5
2.3.3 [2.3] 使用数字技术分享	5
2.3.4 [2.4] 使用数字技术协同	6
2.3.5 [2.5] 使用数字公共服务	6



全民数字素养与技能培训基地

“数字素养框架” (2023)



能力域

「0」 通用数字设备与应用软件

「1」 信息与数据

「2」 沟通与协作

「5」 数字安全

「7」 特定职业相关

「3」 创建数字内容

「4」 创建数字工具

「6」 数字思维与问题解决



全民数字素养与技能培训基地

“数字素养框架” (2023)



0 通用数字设备和应用软件

[0.1] 使用智能电子设备	操作智能手机、平板电脑、智能家电等智能化设备。
[0.2] 使用通用计算机设备	操作通用的个人电脑。
[0.3] 使用常用应用软件	操作常用应用软件，包括办公软件、图形图像工具、通信协同工具等。

1 信息与数据

[1.1] 浏览、搜索和筛选信息数据	在数字环境中浏览各种信息与数据，根据自身需求搜索有用的信息与数据，在多种格式及媒介的信息与数据中导航。
[1.2] 分析、比较和评价信息数据	分析、比较和批判性地评价信息与数据的可信度，对信息和数据进行综合性的分析以得出相对可信的结论。
[1.3] 管理信息与数据	在数字环境中组织、存储和使用信息与数据，必要时对它们做结构化组织、清洗和加工。

2 沟通与协作

[2.1] 管理数字身份	创建和管理自己的一个或多个数字身份，能够保护自己的数字声誉，能够处理自己的数字身份产生的数据。
[2.2] 使用数字技术互动	使用数据技术进行沟通和互动。
[2.3] 使用数字技术分享	使用数字技术与他人分享信息、数据与数字内容，了解引用和注明出处。
[2.4] 使用数字技术协同	使用数字技术实现多人协同，包括对协同的促进和对协同环境中产生信息、数据与数字内容的管理。
[2.5] 使用数字公共服务	定位和使用政府及其他组织提供的数字化公共服务，了解在此过程中保护自身数字权益的方法。
[2.6] 网络礼仪	了解数字环境中使用数字技术与互动的行为规范和具体做法；了解并尊重数字环境中的文化与代际多样性，制定相匹配的沟通策略及规范。

3 创建数字内容

[3.1] 创作数字内容	创作和编辑不同格式与媒体形式的数字内容，使用数字工具表达自己的想法。
[3.2] 数字内容再创作	修改、精炼、整合、改进已有的信息与内容，以创建相关的新内容和新知识。
[3.3] 版权与许可	理解版权与许可应用于数据、信息和数字内容的原理和实践，保证数字内容的创建与传播合规合法。

4 构建数字工具

[4.1] 规划与设计数字工具	理解现实世界和数字世界的需求，设计可实现的、有助于提升数字环境运作效率的软件工具。
[4.2] 创建数字工具	规划和创建计算机系统可理解的指令，实现解决问题或完成任务的软件工具。
[4.3] 管理数字工具	对数字工具的使用者提供持续运营、服务、技术支持和系统维护。

5 数字安全

[5.1] 对数字设备的保护	保护设备与数字内容，理解数字环境中的风险与威胁；了解安全与安保措施，适当考虑可靠性与隐私。
[5.2] 对个人数据与隐私的保护	保护数字环境中的个人数据与隐私；理解使用和分享个人身份信息的安全方式，以保护自己与他人利益不受损害；能够理解数字服务的“隐私政策”，尤其是其个人数据将被如何使用。
[5.3] 对个人健康与福祉的保护	能够在数字技术使用过程中，避免其对身心健康造成威胁；能够在数字环境中保护自己与他人利益不受损害（如网络霸凌）；了解数字技术对社会福祉与社会融入的作用。
[5.5] 对环境的保护	了解数字技术及其使用对环境的影响。

6 数字思维与问题解决

[6.1] 解决技术问题	确认和解决操作设备与使用数字环境过程中的技术问题（从故障检测到解决复杂问题）。
[6.2] 设计技术解决方案	分析问题和评估需求，评估、选择和运用数字工具形成可行的解决方案以满足需求；必要时调整和定制数字环境以满足需求。
[6.3] 创造性地使用数字技术	使用数字工具与技术创造知识、创新流程与产品。
[6.4] 数字素养提升	理解自己需要在哪些方面提升数字素养；能够支持他人提升数字素养；紧跟数字化发展潮流寻求自我发展的机会。
[6.5] 计算思维	将可计算的问题转化为一系列有逻辑顺序的步骤，为人机系统提供解决方案。
[6.6] 数据思维	掌握通过数据分析得到结论的原理、方法、工具及其局限性；能够有意识地设计数据的采集、清洗、统计、分析方案来验证自己的猜想和理论。

高校落地：数字时代的大学生应该具备什么？



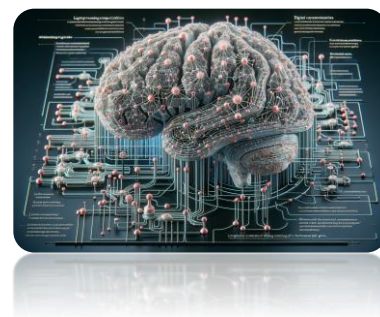
全体
大学生

大学生数字
素养与技能

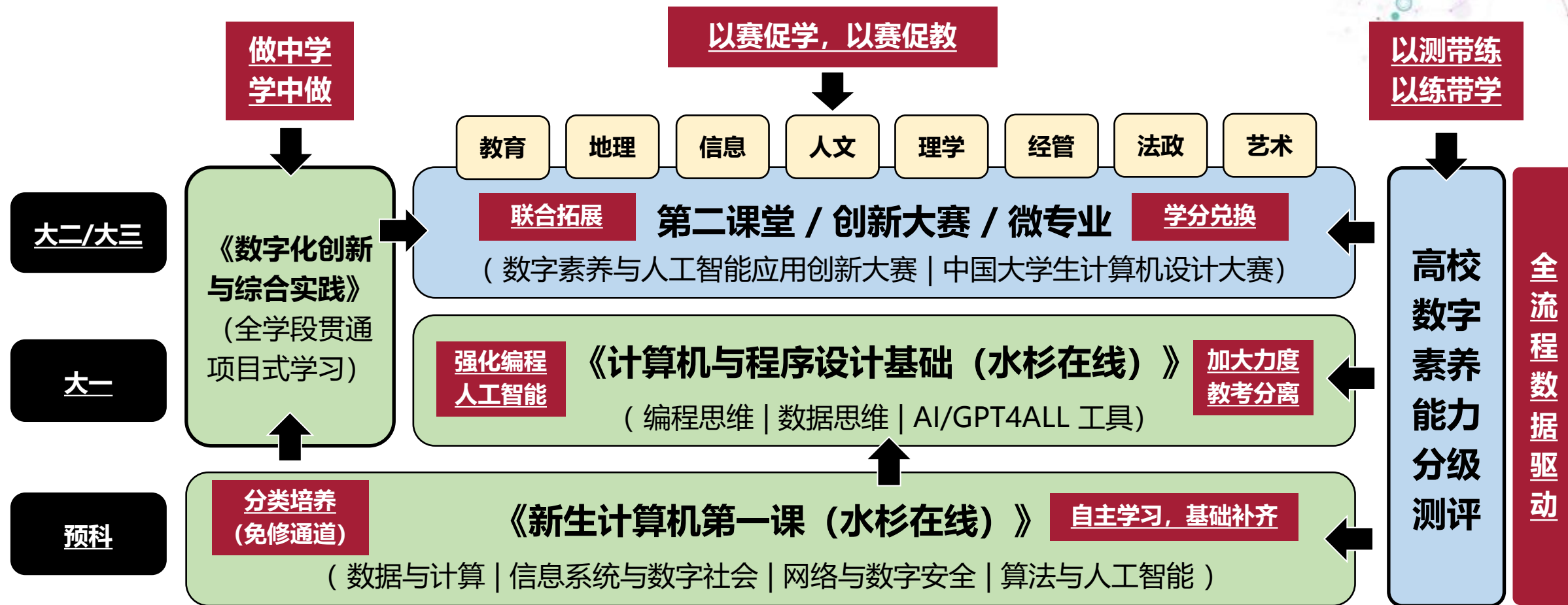
编程思维

数据分析

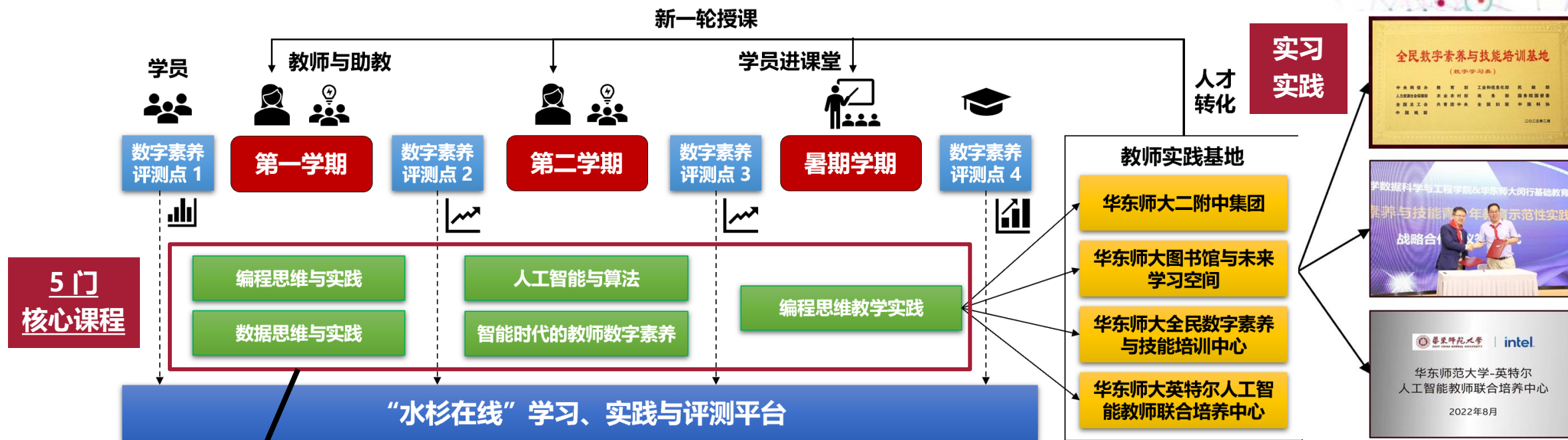
AI / AIGC



举措 1：学生数字素养提升落实方案（计算机公共课）



举措 2: 数字素养教育微专业 (1 个学年 / 10 个学分)



《编程思维与实践》

- 理解编程语言
- 程序的基本结构
- 面向对象的思想与实践
- 聊天机器人设计与实现
- 深入函数和模块
- 函数式编程和字符串
- 数据容器
- 理解数字化世界

《数据思维与实践》

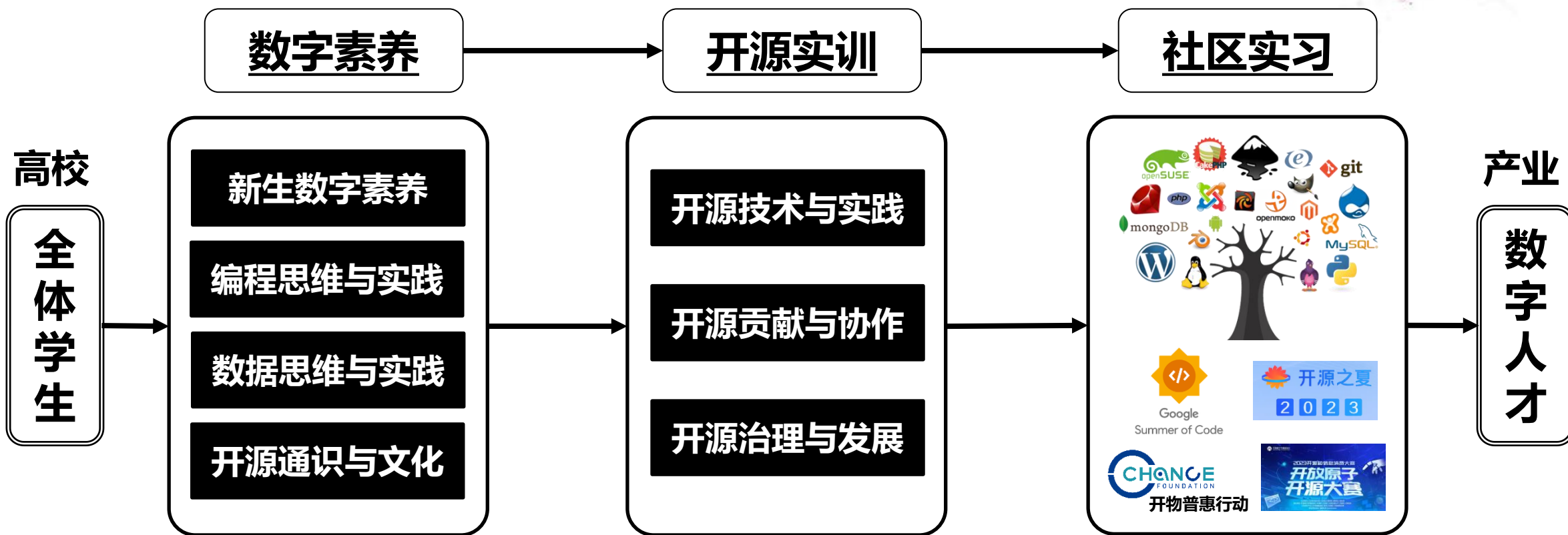
- 数据科学与思维入门
- 数据科学的数学基础
- 数据收集与管理
- 数据探索与预处理
- 数据建模与分析
- 数据科学实践案例

智能时代 II

“This is a knowledge graph. First, a human has to create a topic, and then the AI can generate a series of graphs. The human then chooses and organizes, and AI can only teach them this.”

智能时代 II: 人工智能 → AI 数据 → 人工智能
不能 AI 取代的必须: AI 不知道的领域 (智慧), 理解 AI 不能理解的内容 (解释性), 制作 AI 还不会做的东西 (创造性).

举措 3：开源对教育模式的变革 — 从数字素养到数字人才



举措 4：数字化学习平台



教-学-练-测-创-评

全民社会服务

跃升服务

平台与数据

关键技术

教职工数字素养提升

- 职能部门与院系管理人员
- 各部门数据管理人员
- 数字化卓越辅导员
- 全体教师

学生数字素养提升

- 数字素养第二课堂
- 数字素养能力分级评测
- 数字素养微专业
- 大学计算机基础课程

- 水杉在线
- 全民数字素养与技能培训基地
- 未来学习中心

<h3>数字课本</h3> <ul style="list-style-type: none">标准规范生产工具集成学习环境	<h3>能力测评</h3> <ul style="list-style-type: none">能力图谱智能题库测评与诊断	<h3>GPT4ALL</h3> <ul style="list-style-type: none">API 服务开源定制超算基座
--	---	---

面向“全民数字素养提升”的教育大模型架构



教育场景

AI4Coding

AI4SE

AI4DS

AI4OSS

代码级生成

任务级生成

场景级生成

关键技术

测评与优化

测评规范

测评工具

优化方法

边缘与轻量

AIGC一体机

适配工具链

BigDL 项目

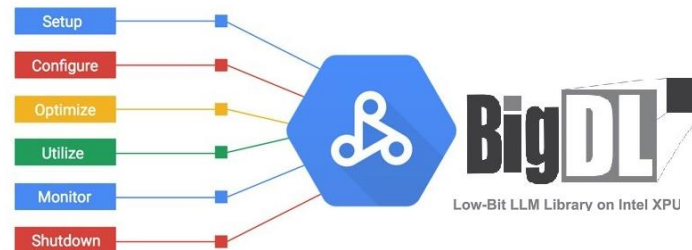
GPT4ALL

教育场景适配

提示词工程

低代码平台

基础设施



GPT4ALL 的研究与进展



教育场景 AI 能力域	教	学	练	测	创	评
基础知识 (AI4DL)	<ul style="list-style-type: none"> 教案生成 出题组卷 知识图谱 	<ul style="list-style-type: none"> 自适应学习 智能助教 	<ul style="list-style-type: none"> 个性化练习 自动纠错 拓展练习 	<ul style="list-style-type: none"> 自动批改 个性化试题 自动打标签 	<ul style="list-style-type: none"> 创意生成 项目推荐 智能查新 	<ul style="list-style-type: none"> 测评大模型 能力画像
编程技能 (AI4Coding)	<ul style="list-style-type: none"> 编程知识图谱 	<ul style="list-style-type: none"> 代码助教 	<ul style="list-style-type: none"> 代码纠错 	<ul style="list-style-type: none"> 代码自动出题 代码自动打标 	—	<ul style="list-style-type: none"> 代码能力评测
工程能力 (AI4SE)	<ul style="list-style-type: none"> 工程知识图谱 	<ul style="list-style-type: none"> DevGPT 	<ul style="list-style-type: none"> AIOps 助教 	<ul style="list-style-type: none"> DevOps 评测 	<ul style="list-style-type: none"> 工程场景生成 工程管理 Agent 	<ul style="list-style-type: none"> 工程能力评测
数据分析 (AI4DS)	<ul style="list-style-type: none"> 数据科学图谱 	<ul style="list-style-type: none"> GPT-DB GPT-GDS 	<ul style="list-style-type: none"> 实训数据生成 分析代码生成 	<ul style="list-style-type: none"> 分析模型评测 SQL/GQL生成 	<ul style="list-style-type: none"> 数据作品辅助 数据场景生成 	<ul style="list-style-type: none"> SQL/GQL评测
开源协作 (AI4OSS)	<ul style="list-style-type: none"> 开源知识图谱 开源项目标签 	—	<ul style="list-style-type: none"> GitOps助教 	<ul style="list-style-type: none"> 开源社区问答 	<ul style="list-style-type: none"> 开源协作机器人 	<ul style="list-style-type: none"> 贡献度评价 开源人才画像
基础设施 (AI4Infra)	<ul style="list-style-type: none"> 基础代码大模型、代码大模型评测、超轻量模型 (BigDL)、教育场景低代码平台、大模型一体机 					

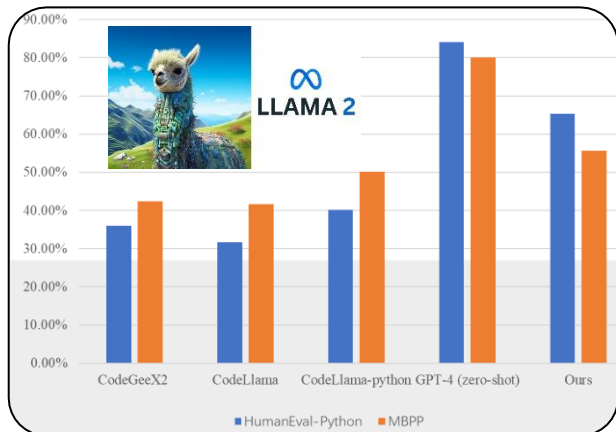
Birmingham 2023
CIKM
The 32nd ACM International Conference on Information and Knowledge Management
October 21-25, 2023
Birmingham, UK

Best Demo Paper Award

Jianing Wang, Nuo Chen, Qiushi Sun, Wenkang Huang, Chengyu Wang & Ming Gao
HugNLP: A Unified and Comprehensive Library for Natural Language Processing

The General Chairs: Mark Lewis, Mark Lewis
The Demo Paper Program Chairs: Jianing Wang, Nuo Chen, Qiushi Sun, Wenkang Huang, Chengyu Wang & Ming Gao

水杉代码大模型



intel digital readiness

大模型实训一体机

DL4ALL
全民数字素养与技能培训基地

给定一个整数数组 nums 和一个整数目标值 target，请在该数组中找出和为目标值 target 的那两个整数，并返回它们的数组下标。你可以假设输入只会对应一个答案。但是，数组中同一个元素在答案里不能重复出现。

输出: [2, 7]

解释: 因为 nums[0] + nums[1] == 9, 返回 [0, 1]。

复杂代码生成

```
python
class Solution:
    def twoSum(self, nums: List[int], target: int) -> List[int]:
        hash_table = {}
        for i, num in enumerate(nums):
            complement = target - num
            if complement in hash_table:
                return [hash_table[complement], i]
            hash_table[num] = i
        return [-1, -1]
```

在测试代码中，我们调用 `fibonacci(10)` 并打印结果。这将输出第 10 个斐波那契数，即 55。

User: 帮我把上一段代码改成 c++ 可以吗
Assistant: 这是一个简单的 c++ 版本的代码:

```
int main() {
    int n;
    cout << "Enter a number: ";
    cin >> n;
    cout << "The Fibonacci number for " << n << " is " << fibonacci(n) << endl;
    return 0;
}

int fibonacci(int n) {
    if (n <= 1)
        return n;
    else
        return fibonacci(n-1) + fibonacci(n-2);
}
```

代码解释

开源人才评价

姓名	GitHub	贡献度
Linus Torvalds	2000+	100%
Guido van Rossum	1000+	95%
Richard Stallman	500+	90%

代码纠错

实施方案:

- 调研和比较传统深度学习方法和大语言模型修复学生代码的效果
- 调研和部署开源大语言模型并在学术代码数据集上进行测评
- 利用提示工程和指令微调，提升大语言模型修复代码的准确率和生成提示信息的质量
- 将大语言模型部署至实训平台，进行效果测试

部分实验结果:

模型	Codefuse-34B	Codefuse-34B	Deepseek-33B	Deepseek-33B
学习方式	zero-shot	one-shot	zero-shot	one-shot
尝试次数	5	5	1	1
平均修复率	93%	85%	73%	71%

融合大语言模型的社区问答系统

服务层: 动态更新, 文档管理, 数据库语言查询, 对话查询, 信息收集, 实时展示, 计算机领域知识检索, PolarDB社区问答

文档管理: 非结构化数据, 同步维护, 历史收集, 向量库存储

大模型问答: 通用知识检索, 外部知识检索, 反馈机制

PolarDB问答

模型	模型参数	bleu4	余弦相似度	最优次数
baichuan	13B	0.033	0.728	46
ChatGLM2	6B	0.033	0.768	25
Qwen	14B	0.08	0.803	5
ChatGPT	175B	0.035	0.644	17

SQL 代码生成

method	easy (163)	medium (298)	hard (114)	extra (90)	all (665)
chatglm2-6b	0.160	0.040	0.061	0.000	0.068
baichuan-13b-chat	0.129	0.067	0.044	0.011	0.071
vicuna-13b-v1.5	0.061	0.017	0.009	0.000	0.024
Qwen-14B-Chat	0.497	0.292	0.263	0.100	0.311

多轮问答

模型	BLEU1	BLEU2	BLEU3	BLEU4
chatglm2-6b	0.143	0.0516	0.0289	0.0198
chatglm3-6b	0.141	0.0532	0.0296	0.0197
vicuna-13b-v1.5	0.143	0.0531	0.0298	0.0196
baichuan-13b-chat	0.157	0.0644	0.0364	0.0240
Qwen-14B-Chat	0.180	0.0815	0.0489	0.0342

计算机常识问答

模型	无知识库查询时间	有知识库查询时间	余弦相似度	BLEU4
chatglm2-6b	13.8838	9.79087	0.6291	0.0104
Qwen-14B-Chat	12.3876	12.5260	0.6659	0.0180
baichuan-13b-chat	16.0946	15.2760	0.6319	0.0085
vicuna-13b-v1.5	10.6538	10.0477	0.6238	0.0084
chatglm3-6b	11.8166	12.9287	0.6225	0.0090

数据统计, 系统迭代, 智能助理, 实时问答

自动出题

实施方案:

- 调研目前主流的大语言模型自动出题的能力，包括GPT-4以及国内外较有代表性的开源大语言模型
- 探索自动出题应用场景更多可能性，包括简单编程题和复杂项目题
- 通过指令微调，提升大语言模型在自动出题任务上的准确性并贴合使用场景，实现多样化特点

部分实验结果:

模型	合理性	涵盖知识点	输入输出样例一致	sample solution 正确	可以减少人工工作量
Qwen	85.45%	97.27%	61.82%	54.55%	67.27%
Ernie	72.60%	90.41%	50.68%	54.79%	56.16%
Spark	78.26%	91.30%	78.26%	93.48%	73.91%
DeepSeek	100%	100%	77.42%	83.87%	93.55%
				98%	96.00%

图谱展示

知识图谱搜索 能力图谱搜索 MATCH res=(n:主题知识)-[]-(知识图谱) RETURN res

关联实体 添加实体 修改实体

知识图谱 (26) 知识单元(1) 知识图谱(13) 知识对象(1) 主题知识(11)

能力图谱 (24) 能力图谱(12) 能力项(1) 能力类(3) 能力点(9)

知识图谱 能力图谱

知识库与知识图谱的构建

基于知识点的教学视频内容检测及自动切片

视频提取知识点检测 - 飞书云云

目标检测智能化训练

简介

上传知识库文件

上传教学视频

展示知识点

视频知识点提取

上传教学视频、返回视频知识点及知识点视频

习题自适应推荐

简介操作指南:

- 1、上传知识库文件
- 2、上传教学视频,支持mp4格式
- 3、返回视频知识点以及知识点视频,可以查看并下载

下一步

基于知识点的学习者画像与自适应学习

知识点自适应学习系统

知识点提取

知识点图谱

习题自适应推荐

知识追踪可视化

存储设备	0.54	0.59	0.64	0.72	0.78	0.82	0.84	0.85	0.82	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.87	0.85
CPU	0.51	0.55	0.6	0.68	0.71	0.75	0.77	0.79	0.79	0.79	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
互联网	0.44	0.44	0.42	0.4	0.38	0.38	0.37	0.37	0.39	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36
计算机网络基础	0.62	0.67	0.74	0.78	0.82	0.84	0.86	0.87	0.86	0.87	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
办公自动化	0.65	0.69	0.81	0.85	0.89	0.91	0.92	0.92	0.94	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
信息系统与数学社会统计	0.59	0.62	0.67	0.71	0.74	0.77	0.78	0.79	0.81	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
操作系统	0.61	0.67	0.73	0.79	0.83	0.87	0.89	0.89	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
计算机系统硬件	0.5	0.54	0.57	0.63	0.68	0.71	0.72	0.73	0.74	0.73	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74
EXCEL数据分析	0.54	0.59	0.64	0.69	0.73	0.77	0.79	0.8	0.78	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81
计算机文化与数学	0.67	0.6	0.62	0.66	0.7	0.72	0.74	0.74	0.73	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75

岗位能力考核及岗位能力画像

数据分析Python

C1. 数据建模、评估和应用

C2. 数据评估与处理

C3. 数据应用

能力域

Java工程岗位能力要求

Java基础知识

数据库技术

网络编程技术

能力域

数据建模、评估和应用

3-3.模型构建与评价

3-2.模型优化

能力项

Java核心技术

3-3.模型构建与评价

数据库技术

网络编程技术

能力项

模型构建与评价

4. 模型构建

1.00

能力项

面向对象编程

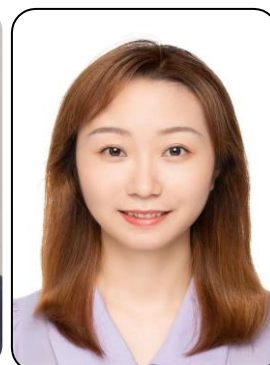
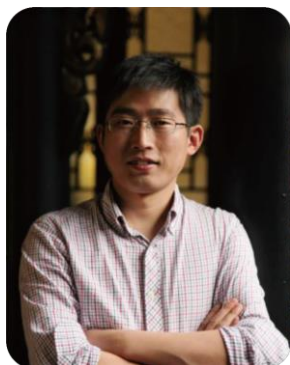
多态

封装和继承

接口实现与使用

测评点

团队与分工



王伟
(教授)
计算教育学
开源数字生态

高明
(教授)
知识图谱
基础大模型

胡文心
(教授级高工)
大模型评测
AI4DL

陆雪松
(副教授)
代码大模型
AI4Coding

兰韵诗
(副教授)
计算语言学
AI4OSS

蒲鹏
(高级工程师)
智能运维
AI4DS

刘垚
(副教授)
高性能计算
AI4Infra

张琰彬
(工程师)
AI DevOps
AI4SE

苏斌
(助理工程师)
数字课本
AI4Infra

相关成果

1. Motivating Open Source Collaborations Through Social Network Evaluation: A Gamification Practice from Alibaba, ICSE 2024. (CCF-A)
2. ErrorCLR: Semantic Error Classification, Localization and Repair for Introductory Programming Assignments. SIGIR 2023. (CCF-A)
3. Uncertainty-Aware Self-Training for Low-Resource Neural Sequence Labeling. AACL 2023. (CCF-A)
4. Tat-Seng Chua: Learning Relation Prototype From Unlabeled Texts for Long-Tail Relation Extraction. TKDE 2023. (CCF-A)
5. A Neural Network Architecture for Program Understanding Inspired by Human Behaviors. ACL 2023. (CCF-A)
6. Plan-and-Solve Prompting: Improving Zero-Shot Chain-of-Thought Reasoning by Large Language Models, ACL 2023. (CCF-A)
7. Improving Zero-shot Visual Question Answering via Large Language Models with Reasoning Question Prompts, MM 2023. (CCF-A)
8. MWPToolkit: An Open-Source Framework for Deep Learning-Based Math Word Problem Solvers. AACL. 2022. (CCF-A)
9. Multi-task Learning for Paraphrase Generation With Keyword Reconstruction. ACL 2022. (CCF-A)
10. Meta-Learning Adversarial Domain Adaptation Network for Few-Shot Text Classification. ACL 2021. (CCF-A)
11. On Disambiguating Authors: Collaboration Network Reconstruction in a Bottom-up Manner. ICDE 2021. (CCF-A)
12. FlaCGEC: A Chinese Grammatical Error Correction Dataset with Fine-grained Linguistic Annotation. EMNLP 2023. (CCF-B)
13. Pass-Tuning: Towards Structure-Aware Parameter-Efficient Tuning for Code Representation Learning. EMNLP 2023. (CCF-B)
14. Evaluating and Enhancing the Robustness of Code Pre-trained Models. EMNLP 2023. (CCF-B)
15. Temporal Autoregressive Matrix Factorization for High-dimensional Time Series prediction of OSS, TNNLS, 2023. (CCF-B)
16. Understanding the Archived Projects on GitHub, SANER 2023. (CCF-B)
17. HugNLP: A Unified and Comprehensive Library for Natural Language Processing. CIKM 2023. (CCF –B, Best Demo Paper Award)
18. Prompting Large Language Models with Chain-of-Thought for Few-Shot Knowledge Base Question Generation, EMNLP 2023. (CCF B)
19. R3 Prompting: Review, Rephrase and Resolve for Chain-of-Thought Reasoning in Large Language Models under Noisy Context, EMNLP 2023. (CCF B)
20. CAT-probing: A Metric-based Approach to Interpret How Pre-trained Models for Programming Language Attend Code Structure. EMNLP 2022. (CCF-B)

总结

《数字素养提升宣言》

完整版

提升探索能力与学习能力比掌握静态知识更重要。
参与计算机科学普及教育是提升数字能力的最佳实践。
有趣又实用的课程是成功关键。
用好大众喜闻乐见的数字教育和媒体工具。
构建全民参与的终身学习社会。

关键词版

Learning to learn
Computer Science for all
Next generation digital courses
e-Learning on any platform
Everyone and everywhere