

2025 年上海市高等学校信息技术水平考试试卷

四级 人工智能 II 类 (A 场)

(本试卷考试时间 150 分钟)

一、单选题 (本大题 20 道小题 , 每小题 1 分 , 共 20 分), 从下面题目给出的 A、B、C、D 四个可供选择的答案中选择一个正确答案。

1. _____ 技术主要用于分析词之间的依赖关系。

- A. TF-IDF
- B. 依存句法分析
- C. Transformer
- D. SVM

2. Word2Vec 中, CBOW 模型的基本思想是_____。

- A. 预测上下文
- B. 预测中心词
- C. 最大化条件概率
- D. 训练图神经网络

3. 以模拟生物神经网络、通过权重学习来实现认知功能, 并在深度学习中得到快速发展的流派是_____。

- A. 贝叶斯主义流派
- B. 连接主义流派
- C. 符号主义流派
- D. 强化学习流派

4. 从系统学角度看, 智能计算系统的“定义”中最关键的要素组合是_____。

- A. 软件、硬件和电力供应
- B. 感知、认知与执行三大功能模块
- C. 数据、算法与网络协议
- D. 芯片、操作系统与数据库

5. 大数定律主要说明_____现象。

- A. 样本平均值在样本量足够大时趋近于总体期望
- B. 任何事件的概率都会随着实验次数增加而减少
- C. 两个独立事件的联合概率等于概率之和
- D. 样本方差随着样本量增大必定为零

6. _____ 属于数据管理的关键任务。

- A. 编写深度学习代码
- B. 清洗和标注数据
- C. 购买 GPU 服务器
- D. 设计神经网络结构

7. _____是算力优化的常见方法。
- A. 增加数据标注人员
 - B. 使用分布式计算
 - C. 减少特征数量
 - D. 忽略缺失值
8. 在 Transformer 架构的 Decoder 中, Multi-Head Attention 的 Q (Query)、K (Key)、V (Value) 来自_____。
- A. Q、K、V 均来自 encoder
 - B. Q、K 来自 encoder, V 来自 decoder
 - C. Q 来自 decoder, K、V 来自 encoder
 - D. Q 来自 encoder, K、V 来自 decoder
9. _____属于无监督学习的典型任务。
- A. 图像分类
 - B. 客户分群
 - C. 房价预测
 - D. 垃圾邮件识别
10. _____是数据预处理的常见步骤。
- A. 模型部署
 - B. 缺失值填充
 - C. 购买服务器
 - D. 编写 API 接口
11. 性能评估指标中, F1 值是_____的调和平均数。
- A. 准确率与召回率
 - B. 精确率与召回率
 - C. 方差与标准差
 - D. 训练损失与验证损失
12. 神经网络反向传播的作用是_____。
- A. 计算输出
 - B. 计算梯度, 指导参数更新
 - C. 初始化参数
 - D. 增加隐藏层
13. 深度学习模型训练中, 过拟合的表现_____。
- A. 训练误差高, 测试误差高
 - B. 训练误差低, 测试误差高
 - C. 训练误差与测试误差均低
 - D. 训练误差与测试误差均高
14. 在稀疏三维重建中, 通常使用_____算法进行特征匹配。

- A. PCA
- B. SIFT
- C. GAN
- D. RNN

15. 在词法分析中，词性标注的主要目的是_____。

- A. 把词变成数字
- B. 为每个词分配一个语义标签
- C. 判断每个词的语法类别
- D. 检测实体边界

16. RNN 相对于传统 n-gram 模型的主要优势是_____。

- A. 可训练词性
- B. 捕捉长期依赖
- C. 提高语法正确率
- D. 节省内存

17. 上海人工智能产业链中，常见的“边缘计算”主要指_____。

- A. 在数据中心进行大规模 AI 模型训练
- B. 在数据产生的终端或附近进行数据处理和分析
- C. 专注于网络安全的 AI 解决方案
- D. 纯软件层面的云端 AI 服务

18. 下列应用中，_____是具身智能的一个典型应用场景。

- A. OCR 合同文本识别
- B. 自动驾驶清扫机器人
- C. 云端数据加密
- D. 大规模文本摘要

19. 以下关于提示（prompt）作用的描述，正确的是_____。

- A. 给模型的输出添加随机性
- B. 使模型忽略特定任务
- C. 提示模型针对某一特定任务该如何输出
- D. 扩大模型的词表

20. 在昇思 MindSpore 中，可以通过_____方式将 Tensor 变量转换为 Numpy 变量。

- A. Tensor()
- B. Tensor.asnumpy()
- C. Tensor.astype()
- D. Tensor.assign_value()

二、多选题（本大题 10 道小题，每小题 2 分，共 20 分），从下面题目给出的 A、B、C、D 四个可供选择的答案中选择所有正确答案。

1. _____是回归算法。
- A. 线性回归
 - B. 逻辑回归
 - C. 决策树
 - D. K-Means
2. _____是自然语言处理中常见的序列标注任务。
- A. 命名实体识别
 - B. 依存句法分析
 - C. 语音识别
 - D. 文本摘要
3. _____是多模态融合中的模态。
- A. 文本
 - B. 语音
 - C. 视频
 - D. 图片
4. 防止过拟合的有效方法有_____。
- A. 增加数据
 - B. 正则化
 - C. 减少模型复杂度
 - D. 增加训练轮数
5. _____方法属于机器学习在自然语言处理中的数据增强方法。
- A. 同义词替换
 - B. 人工标注
 - C. 词语模糊
 - D. 随机删除
6. _____属于 Transformer 架构的核心机制。
- A. 自注意力机制 (Self-Attention)
 - B. 卷积神经网络 (CNN)
 - C. 多头注意力 (Multi-Head Attention)
 - D. 位置编码 (Positional Encoding)
7. 在自然语言处理任务中，语料库的主要作用包括_____。
- A. 提供模型训练数据
 - B. 作为算法评估的测试集
 - C. 帮助语言学研究
 - D. 直接替代神经网络结构

8. 在零售行业中，人工智能可用于_____场景。

- A. 智能推荐系统
- B. 库存预测与管理
- C. 虚拟试衣
- D. 自动驾驶车辆调度

9. 关于人工智能伦理，_____原则是普遍认可的。

- A. 公平与无歧视
- B. 透明性和可解释性
- C. 自动化取代所有人类工作
- D. 责任和问责制

10. 昇思 MindSpore 支持对_____模块进行自定义。

- A. 评价指标
- B. 损失函数
- C. 优化器
- D. 神经网络层

三、是非题（本大题 10 道小题，每小题 1 分，共 10 分）。

1. 生成式人工智能的三大要素是数据、算法和算力。

2. 强化学习中，智能体的目标是最大化累计奖励。

3. “具身智能”认为智能行为不仅依赖大脑计算，还受身体结构和与环境交互方式的共同影响。

4. 在经典 DNN-HMM 语音识别框架中，DNN 主要负责估计声学状态转移概率，而 HMM 负责估计每帧输入与隐状态的发射概率。

5. 在自然语言处理中，词嵌入(Word Embedding)的目的是将词汇映射到实数向量空间。

6. 在数据并行训练中，如果所有工作节点使用相同的模型结构并且拥有各自独立的数据子集，那么每轮迭代后必须同步所有模型参数，以保证全局模型的一致性。

7. 在大规模生成模型中，若预训练数据存在明显的偏见或不当内容，这些偏差一般会在微调阶段被自动消除，不会影响最终模型输出。

8. 在人工智能系统开发中，若训练数据公开可得（如网络爬取），开发者便无需考虑数据主体的知情同意与使用目的合法性。

9.昇思 MindSpore 作为深度学习框架，下接 CANN 异构计算架构的硬件使能层，上接如 MindSpeed、MindIE 的应用使能层。

10.昇腾 AI 处理器主要面向人工智能计算场景，支持深度学习训练和推理任务。

四、操作题

以下案例应用题题目请在文件“C:\KS\人工智能-答题纸E卷.docx”中作答！

（一）简答题(本大题包含2小题，每题5分，共10分)

1.随着人工智能向大模型发展，今年DeepSeek、Qwen、InternLM等开源项目引领了行业的快速发展，请写出大模型生态里面的强化学习、推理引擎、训练框架、Agent等5个优质开源项目。（5分）

请在答题纸作答！此处答题一律无效！

2.今年年初DeepSeek的爆火，除了领先的技术以外，还举行了开源周，公开了大量的核心技术，促进了整个行业的突破。持续学习和读懂前沿技术是行业从业的必备能力，开源周有一篇关于DeepEP的介绍（见C:\素材\），请结合附件写出MoE的功能和技术亮点。（5分）

请在答题纸作答！此处答题一律无效！

（二）基础操作题(本大题包含2小题，每题15分，共30分)

1.在业务实际操作中，需要对环境和服务器进行部署，例如安装基本的PyTorch、MindSpore等AI框架，安装基本的sklearn、transformer、Matplotlib等库包。请回答如下问题：

（1）请完成如下两个任务，填写在答题纸中：（5分）

- ①使用pip命令安装最新版的sklearn库；
- ②检查sklearn库的版本号。

请在答题纸作答！此处答题一律无效！

（2）应某企业应用需求，需要在其公司新购置安装Windows 10操作系统的服务器上安装Qwen，目前已安装了Ollama，请接着完成如下安装Qwen（Qwen3:1.7b）的任务，该任务包括三个步骤，请将命令写在答题纸中。（5分）

- ①创建和安装Qwen模型；
- ②运行模型；
- ③检查包括Qwen在内的各个模型版本。

请在答题纸作答！此处答题一律无效！

（3）在Linux操作系统环境中,日常需要完成Docker容器的日志排查，AI训练脚本的运行等工作。请撰写代码命令满足如下要求。（5分）

场景：AI 训练脚本运行报错 Permission denied，模型文件 /data/models/resnet50.pth 无法读取。

要求：

- ①检查文件权限（仅需显示权限位，如 -rw-r--r--）；
- ②为 ai_user 用户组添加读写权限（不改变其他用户权限）；
- ③检查权限修改结果（输出修改后权限）。

请在答题纸作答！此处答题一律无效！

2. 接下来公司研发部要在服务器上开展AI的研发，例如利用深度神经网络对图像进行训练和分类预测，或者利用预训练模型进行分类预测等等，请补充代码完成以下任务。（15分，每空3分）

（1）在给定的一个预训练ResNet50模型上，对输入的图像（image.jpg）进行目标分类，并输出预测类别。以下为核心代码，请补充空白处。

```
from torchvision import models, transforms
from PIL import Image
import torch

model = models.resnet50(pretrained=True)
model.eval()

transform = transforms.Compose([
    _____ ① _____, #图像大小统一为 32x32
    transforms.CenterCrop(224),
    _____ ② _____, #图像转为Tensor
    transforms.Normalize(mean=[0.485, 0.456, 0.406],
                          std=[0.229, 0.224, 0.225])
])

img = Image.open("_____ ③ _____") #打开待处理图像
img_t = transform(img).unsqueeze(0)

with torch.no_grad():
    output = model(img_t)
    _, predicted = torch.max(output, 1)

print(f"Predicted class index: {predicted.item()}")
```

请在答题纸作答！此处答题一律无效！

（2）使用 Hugging Face Transformers 库加载 bert-base-chinese 模型，实现中文句子情感分类（正向/负向）。以下是关键流程和代码，请补充空白处。

关键流程：

- (1) 加载分词器与模型
- (2) 对输入文本进行分词编码
- (3) 前向推理并获取 logits
- (4) 取最大概率类别作为预测结果

关键代码：

```
from transformers import BertTokenizer, BertForSequenceClassification
import torch
```

```
tokenizer = BertTokenizer.from_pretrained('bert-base-chinese')
model = BertForSequenceClassification.from_pretrained('bert-base-chinese',
num_labels=2)
```

```
inputs = tokenizer("这家餐厅真不错", return_tensors="pt")
outputs = model(inputs)
pred = _____④_____ (outputs.logits, dim=1)
print(pred. _____⑤_____ ) # 0-负向, 1-正向
```

请在答题纸作答！此处答题一律无效！

(三) 实践应用题(本大题包含2小题，每题15分，共30分)

1.近年来，随着人工智能技术在医疗健康领域的快速发展，AI诊断辅助系统已成为提升诊疗效率与准确性的关键工具。某三甲医院计划部署一套AI诊断辅助系统，旨在实现医学影像智能分析、电子病历自然语言处理及跨科室数据安全共享，以支持医生进行精准诊断。该系统需处理CT、MRI等医学影像，理解病历文本内容，并在保障患者隐私的前提下整合多源医疗数据。

你作为项目组的AI架构师，请结合医疗AI的技术体系与实际落地挑战，完成以下问题。

- (1) 请列举支撑“AI诊断辅助系统”所需的核心AI技术（不少于3项），并简要说明其作用。（5分）

请在答题纸作答！此处答题一律无效！

- (2) 请描述该系统的辅助诊断流程（如：从接收患者影像和病历到生成诊断建议的全过程），并标出关键AI模块或步骤。（5分）

请在答题纸作答！此处答题一律无效！

- (3) 该系统在肺部CT影像分析中出现“误诊肺结节为正常肺部”问题，请分析原因并提出相应的改进策略。（5分）

请在答题纸作答！此处答题一律无效！

2. 随着普通话的普及，方言正在逐步消失，但是目前老年人在生活中大多数时间仍使用方言。为了方言的保护以为在政务、公共服务等场景上为老人提供更贴心的服务，各地都积极打造独特的方言库和相关数字人。

上海拥有独特的沪语体系，上海大学发布沪语大模型“小沪2.0”，以助力方言保护和政务服务，同时上海大学还将发布“基于沪语模型的创新创业大赛”的项目，鼓励更多有志青年加入到这个项目中来，实现成果转化。

下面我们将利用“小沪2.0”这个基础模型以及RAG技术对养老院的陪伴系统进行改造。基于该背景，完成以下题目。（15分）

（1）请设计一套基于陪伴系统完整的RAG（检索增强生成）系统搭建流程，需重点考虑沪语特性（如方言词汇、语法差异、语音文本转换），明确各环节的具体方案。（5分）

请在答题纸作答！此处答题一律无效！

（2）请写出沪语文本预处理与向量入库的具体操作。（需包含方言词汇预处理、文本分块、沪语适配的 Embedding 模型调用）（5分）

请在答题纸作答！此处答题一律无效！

（3）为了服务的便捷性，需针对系统响应时间和准确率进行提升，提出 3 项针对性优化策略，并说明其原理。（5分）

请在答题纸作答！此处答题一律无效！

（四）场景设计与行业应用题(本大题包含1小题，每题30分，共30分)

2025世界人工智能大会暨人工智能全球治理高级别会议以“智能时代 同球共济”为主题，设置会议论坛、展览展示、赛事评奖、应用体验、创新孵化五大板块，全景呈现AI技术前沿、产业趋势与全球治理的最新实践。大会聚焦国际化、高端化、年轻化、专业化，着力打造“产业创新策源地、垂类应用示范地、产业生态聚集地、创新创业首选地、治理合作先行地”五大高地，推动大会迈向更高专业度、国际化和创新性。

那随着我们的步伐来探索AI时代的奥秘！

1. 首先来到展区，我们首先看到昇腾384超节点Atlas 900 A3 SuperPoD，其通过三大技术优势，入选WAIC 2025镇馆之宝。它通过高速互联总线，突破互联瓶颈，让超节点像一台计算机一样工作。相比传统集群，主要有以下3大优势：

① 超大带宽：超节点内任意两个AI处理器之间通信带宽，相较于传统架构提升15倍，超节点内单跳时延降低10倍。

② 超低时延：昇腾超节点支持全局内存统一编址，具备更高效的内存语义通信能力。通过更低时延指令级内存语义通信，可满足大模型训练/推理中的小包通信需求，提升专家网络小包数据传输及离散随机访存通信效率。昇腾384超节点Atlas 900 SuperPoD是业界唯一突破Decode时延15ms的方案，满足实时深度思考下的用户体验需求。

③ 超强性能：经过实际测试，在昇腾超节点集群上，LLaMA3等千亿稠密模型训练性能可达传统集群的2.5倍以上。在通信占比更高的Qwen、DeepSeek等多模态、MoE模型上，可以达

到3倍以上的提升。

在大模型的分布式训练中，带宽性能一直是一个难点。数据并行（Data Parallelism）和模型并行（Model Parallelism）是大模型分布式训练的两种核心策略，请写出两者在通信需求上的核心差异。（5分）

请在答题纸作答！此处答题一律无效！

2. 下面我们来到商汤展区。现场有大量可爱的毛绒玩具。相比之前的玩具，依托大模型的能力，让其对话更活灵活现，使孩子多个贴心的玩伴。该产品应用到了大模型垂直领域微调，请结合儿童陪伴玩具的对话模型优化（如儿童口语化、短语化较多等特点），写出模型微调流程和注意点。（5分）

请在答题纸作答！此处答题一律无效！

3. 随着参观的持续，发现现场除了多种多样的大模型应用以外，今年开源的氛围大幅提升，除了DeepSeek、Qwen、InternLM等大模型开源以外，国内AI框架--昇思MindSpore持续开源开放，兼容主流推理引擎vLLM、SGLang等三方生态，也联手启智、书生、魔乐等开源社区纷纷打造多元化活动，为开发者打造最佳贡献成长之路。其中，推理引擎vLLM作为今年开源社区最佳的项目之一，其依托独特的技术，极大的增加了推理性能。请写出vLLM的PagedAttention的核心思想与内存优化机制。（5分）

请在答题纸作答！此处答题一律无效！

4. 现场除了大模型推理以外，MoE模型的比例远超去年同期。请对比DeepSeek-MoE的“细粒度专家”与Mixtral的“标准专家”之间的设计差异，并说明各自的适用场景。（5分）

请在答题纸作答！此处答题一律无效！

5. 来到展览馆二楼，现场聚集了上百家具身智能的企业，如智元、宇树等企业。请问如何设计RL框架，使灵巧手完成“抓取→扫码→放置→打结”的序列任务？（5分）

请在答题纸作答！此处答题一律无效！

6. 离开世博展览馆，我们来到科学计算（AI4S）相关论坛。无论是去年诺奖的归属还是近年来基于昇思MindSpore打造的商飞“东方翼风”大模型、南方电网“驭电”大模型荣获WAIC SAIL奖，都表明在通往AGI的路线上离不开AI4S。请说出AI4S大模型和传统NLP大模型的区别。（5分）

请在答题纸作答！此处答题一律无效！

（五）附加测试题

智能体（Agent）具备自主决策能力、能与环境交互并完成任务的AI系统，是大模型产业落地重要途径之一，掌握其搭建能力也是考量大模型素养提升的重要标志。因此特增设附

加测试。

上海，作为全球文化枢纽，每年汇聚海量旅游活动、国际级演唱会及盛大漫展（如上海国际漫展），吸引数百万游客。这些精彩纷呈的文化盛宴，生动彰显城市创新活力与包容精神，使其成为世界级文化中心地位和中外文化交流的璀璨窗口，持续引领全球文化潮流。

漫展等二次元活动广受青少年的热爱。需要您通过Agent Flow开发平台（智算网址：hpc5000.shu.edu.cn）所提供的平台搭建智能体，可以实现一句“帮我看看最近2个月上海二次元活动的情况”，获得活动信息，报名方式，并生成交通出行提示，天气预告等信息。帮助二次元爱好者更好的参与活动。

具体要求：

1. 创建智能体组件：

新建一个名为“漫展助手”的智能体应用。根据你的分析，在应用中创建出对应的各个组件节点（如：语言交互、交通规划、行程输出、天气提醒等）；

2. 配置各组件的提示词；

3. 连接工作流：

4. 测试与验证：

完成工作流搭建后，使用以下用户请求进行测试：

在会话中输入“帮我看看最近2个月上海二次元活动的情况”。

在测试时，智能体从上述网址中成功获得活动信息，报名方式，并生成交通出行提示，天气预告等信息。

5. 需提交的成果

建立Agent.docx文件，包含两张清晰的屏幕截图，要求如下：

① 展示你搭建完成的最终工作流的截图。

② 测试后返回信息的截图。

此文件需要存放在C:\KS\目录下。