人工智能B题库

第一章 人工智能技术概论（15题）

一．选择

1. 下列哪项不属于典型的人工智能研究领域

生物遗传工程

2.哪项技术是实现人工智能的关键技术，它使计算机能够从数据中学习并做出决策或预测 深度学习

3.人工智能的研究始于1956年

4.电子表格不是人工智能的主要技术

5.纺织制造领域未直接提及作为AI的主要应用方向

6.中间神经元网络层次的数量较多最能体现“深度”在深度学习中的含义

7.如果一台机器能够让超过30%的测试者误认为他是人类，则认为这台机器通过了图灵测试

8.人工智能的”弱人工智能”指的是：专门针对某一任务的智能系统

9.能够执行任何算法，解决任何可计算问题的计算系统 最精准的定义了“图灵完备”

10.相较于弱人工智能，强人工智能最本质的区别在于：是否具有跨领域推理和解决问题的能力

二．填空

1.机器学习是一种人工智能技术，它使计算机能够在不进行明确编程的情况下从数据中学习并改进其任务执行能力

2.在深度学习架构中，卷积神经网络尤其擅长处理图像和数据视频，通过卷积层能够捕捉局部特征并减少参数量，从而提高学习效率和准确性。

3.自然语言处理是人工智能领域的一个重要分支，它致力于让计算机能够理解、解释和生成人类的自然语言，如中文和英语

4.在图灵测试场景中，如果一位评判者无法准确区分与之交流的是人还是机器超过30%的时间，则认为该机器通过了测试

5.与弱人工智能相比，强人工智能不仅限于特定应用领域，而是拥有跨领域的知识应用、学习、推理以及自我改进能力

6. 人工智能的第一次发展浪潮始于20世纪50年代，核心方法是符号主义，侧重于使用符号逻辑来模拟人类的理性思考过程。

7.人工智能领域三大主要学派：链接主义、符号主义和行为主义

8.随着技术的不断进步，人工智能的发展趋势之一是向通用人工智能发展，旨在使AI系统能够处理更多样、更复杂的任务，而不仅仅是单一领域的应用

9.自动驾驶汽车是人工智能技术在交通出行领域的革命性应用

10.在医疗健康领域，人工智能通过图像识别技术辅助医生进行疾病诊断，特别是医学影像分析方面展现了高精度和高效性

第二章 机器学习与深度学习（20题）

1. 选择
2. 关于机器学习与深度学习机器学习像是一个大家庭，而深度学习是这个家族中一个特别聪明的年轻成员是正确的
3. 监督学习需要标签，无监督学习不需要标签属于监督学习和无监督学习的区别
4. 机器学习按照学习模式的不同，可分为监督学习、无监督学习和强化学习
5. 现有一定数量的房屋数据，包括房屋的面积、卧室的数量、地理位置等信息以及对应的房价。线性回归方法可以更好地帮助我们预测一些新的房屋的房价
6. 在使用机器学习来教电脑识别苹果和橙子的图片时，编写一系列详细的规则告诉电脑，如果图片是红色且圆形，那很可能是苹果 最不适合用来处理这类分类任务
7. 在使用身的学习训练一个识别猫和狗的图像模型时，为了提高模型的泛化能力，数据科学家可能会采用图像数据小增强功能。复制图像，并重新命名图像文件操作不属于常用的图像数据增强方法
8. 在无监督学习中，研究人员尝试让计算机从未标记的数据中自动发现模式，场景将社交网络上的用户按照兴趣相似度分组最适合应用无监督学习方法
9. 假设你正在训练一个机器学习模型来预测房价，在训练集上表现非常差，但在测试集上表现非常好情况表明你的模型可能出现“过拟合”问题
10. 线性回归、逻辑回归、K-平均算法属于常见的机器学习算法
11. 逻辑回归算法常用于分类问题
12. 填空
13. 监督学习的经典算法可以分为分类、回归两类
14. 无监督学习经典算法包括聚类、降维、异常检测等
15. 卷积神经网络因其在处理图像数据方面的高效性而闻名，其主要特点包括卷积层、池化层和全连层
16. 无监督学习时一种机器学习方法，他在没有明确标签的情况下分析数据，目的是发现数据中的隐藏结构或模式
17. 深度学习的主要流程是数据采集、数据预处理、模型训练与测试、模型评估、新样本预测
18. 数据主要分成结构化数据、非结构化数据两类，其中结构化数据有固定的格式，像我们电脑里整齐排列的Excel表格，里面有各式各样的数字、日期和分类信息，而非结构化数据没有固定的格式，就像是一片文本、一张图像、一段音频等
19. 数据预处理包括数据清理、归一化/标准化、数据增强、数据分割等方法
20. 过拟合和欠拟合是训练模型过程中常见的问题，其中过拟合常常发生在模型太复杂的情况下，欠拟合常常发生在模型过于简单的情况下
21. 深度学习的常见模型 包括卷积神经网络、循环神经网络、Transformer、生成对抗网络、深度强化网络
22. 梯度下降是最常用的优化算法之一，用于最小化深度学习网络的损失函数

第三章 计算机视觉（10题）

1. 计算机视觉的主要任务不包括语音识别
2. 计算机视觉的早期研究主要集中在基本图像处理
3. 图像分类的一个典型应用是人脸识别
4. 一般的目标检测算法，其主要输出内容是目标的边界框
5. 图像分割的目的是将图像分割成多个有意义的部分
6. 图像恢复技术通常用于减少图像噪音、增强图像细节、修复损坏的图像

7. 图像检索的目标是找到与查询图像相似的图像

8. 视频里接技术的一项常见应用是目标跟踪

9. 人体行为识别的一个应用是安全监控

10. 异构多模态视觉融合的一种应用是体感游戏

1. 填空
2. 计算机视觉的目的是让机器具备像人类一样的视觉能力
3. 计算机视觉的发展历程中，图像处理是最早期的研究方法
4. 举出三个图像恢复技术的应用：低光照恢复、图像修复、图像去噪
5. 目标检测的关键问题是找到图像中目标的位置
6. 视频数据处理的首要任务是将视频转换为帧序列
7. 多模态数据类型包括同构数据和异构数据两类
8. 图像检索是根据图象内容搜索相似图像，例如寻找同款式的衣服
9. 图像数据处理的一个关键步骤是特征提取，用于从图像中提取关键信息
10. 图像分割技术广泛应用于医学图像分析，如肿瘤检测
11. 同构数据指的是来源于相同传感器的数据，如多张相机的照片

第五章 智能语音（2题）

一．选择

1.计算机视觉不属于智能语音技术的主要组成部分

2.关于语音信号采集采样频率越低，采样的间隔较短，音频的损失也就越小的说法是不正确的

3.说话人辨认是一种一对多的分类问题是正确的

4.关于被动声纹识别的描述不正确的是被动声纹在进行声纹注册时，系统要求说话人读出固定的文本内容

5.关于声纹识别不正确的是环境噪音对声纹识别无干扰

6.语音识别与声纹识别时相同的概念是不正确的

7.关于语音识别描述不正确的是语音识别可以用于识别说话人是谁

8. 自动配音不属于语音识别的应用

9.百度APP小说听书功能和语音合成智能语音技术有关

10.声音克隆时需要辨认说话人身份，该技术和声音克隆无关

二．填空

1.语音信号采集过程中，模拟语音信号经过采样和量化后变成整型数值

2.声纹识别或说话人识别是基于声音判别说话人身份的一种生物特征识别技术

3.根据不同的任务和应用场景，声纹识别可分为说话人辨认和声纹辨认

4.语音合成，也称为文本转语音，是一种将文本转化为自然流畅的语音的技术

5.声音克隆/语音克隆/声音复刻是一种模仿人类声音的技术

第六章 生成式人工智能（9题）

一．选择

1.AIGC代表：Artificial Intelligence Generated Content

2.智能涌现现象通常发生在当AI模型在实际应用中表现出比预期更好的能力时情况下

3.以文生文技术的一个主要优势是减少人工干预

4.在使用生成式AI写故事时，生成故事大纲步骤是最关键的

5.生成挨骂的一个实际应用场景是为应用程序生成功能代码

6.VoiceEngine在日常生活中的一个实际应用场景是生成语音助手的回答

7.Sora在生成视频中的主要优势是减少视频编辑时间

8.AIGC领域中的主要挑战包括哪些：数据隐私和安全问题、缺乏用户反馈、高昂的硬件成本

9.以文生音技术在教育领域的一个重要应用是语音朗读教学材料

10生成视频的AI技术可能带来哪些新的商业机会：增强现实广告、自动化客服系统、虚拟主播

第七章 人工智能伦理（4题）

一．选择

1.过度挖掘社交媒体资源，会带来大数据的环境下对个人隐私的侵犯

2.人工智能的理论和技术背后确实存在一些难以预知的后果和难以控制的风险。尤其在 通用人工智能技术加速进步的情况下，未来的风险难以估测

3.Vinge预测，在未来30年内，人类将拥有创造超级人工智能的能力。他写到:“不久之后，人类时代就会结束”。这是一个警告，他是最早的人工智能威胁论提出者，和现在的霍金、马斯克所担心的一样。

二．判断

1.由于大数据只能体现出数量而不能进行质量上的判断，所以在采样并不完整的情况下给出的结论未必准确。 （√）

2.机器伦理学是指造人工智能的主体自身的道德。（√）

3.对人类自身的认知架构研究，有助于人工智能伦理学的研究。（√）

三．简答题

1.人工智能在医疗领域的应用是否违背了患者的隐私权？

2.人工智能技术是否会导致大规模失业，从而加剧社会不平等？

3.大模型生成的内容是否受版权保护？

4.人工智能算法的偏见和歧视如何影响社会公正和平等？

简答题未提供正确答案

分数分布（仅供参考）：选择题共60题，90分，（具体分布参考单元后附加括号）

填空题5题，共10分（无具体分布）

学习通习题补充

一．随堂练习

1. 在有监督的学习中，模型通过标记好的训练数据集来进行学习
2. 在强化学习中，智能体通过与环境的交互与奖励信号来优化其行为
3. 无监督学习中的聚类算法试图将数据点分成不同的组，这些组是基于数据点之间的相似性定义的
4. 线性回归算法主要用于哪些类型的预测 连续数值预测
5. 随机森林算法通过什么方式提高模型的准确性和稳定性

多个决策树的预测结果结合

1. 在支持向量机中，被用来最大化不同类别间隔的数据点称什么

支持向量

1. 支持向量机的核心思想是什么

最大化不同类别之间的间隔

1. 神经网络的输入层类似于什么

视网膜

1. 深度学习的主要优势之一是什么

减少对人工特征工程的依赖

1. 深度学习模型的训练通常需要什么

大量标记数据

1. 神经网络模型的灵感来源于什么

人脑神经系统

1. 在神经网络中，哪一层负责深入分析输入数据的特征

隐藏层

1. KNN算法的核心思想是什么

找到特征空间中最临近的K个样本，并根据这些样本的类别来决定新数据的类别

1. 回归问题属于哪种类型的机器学习？

有监督学习

1. 线性回归模型的目的是什么？

预测连续数值

1. 线性回归模型的优点是什么

数学表达简单，易于实现，容易理解

1. 随机森林回归在处理拿着总关系时表现更为出色

非线性关系

1. 在数据预处理中，将数据集从小到大排序后位于中间位置的值，用于描述数据集中趋势，对异常值不敏感的统计量是？

中位数

1. 下列哪个统计量是样本数据值与均值之间差异的平方的平均值，用来描述数据的离散程度，并且对异常值敏感

均方差

1. 在特征编码中，那种编码方式将每个类别转化为一个单独的二进制特征，该特征的值为1/0，表示该样本是否属于该类别，适用于无需的类别特征

独热编码

1. 在特征缩放中，哪种方法是根据特征最大值和最小值将特征缩放到【0,1】区间上，通过改变数据的尺度或单位，消除量纲差异

归一化

22.标准化的特点是什么

将数据转化为具有均值为 0，标准差为 1 的分布

23.在数据预处理中，哪一种方法可以减少极端值的影响，使分布更加对称

对数化

24.在数据预处理中，哪一种统计量是数据集中出现次数最多的值

众数

25.哪一个指标是衡量模型拟合优度的统计量，其值范围在0到1之间，越接近1表示模型的解释能力越强，拟合效果越好

决定性系数

1. 测一测

1.20世纪50年代初，用于判断一台机器是否判断具备思维测试的是图灵测试

2.在人工智能发展史商，第一阶段为20世纪50年代-80年代，发展遇到的瓶颈是随着计算任务的复杂性不断加大，计算能力不断满足需求

3.不属于人工智能的主要分支是虚拟现实

4.不是人工智能应用领域的是传统手工编织工艺

5.人工智能领域的“弱人工智能”和“强人工智能”之间的主要区别在于能够执行的任务范围与智能水平

6.计算机在某些领域的感知能力已越来越接近人类，主要得益于运算能力、数据、算法三大要素的发展

7.关于人工智能、机器学习、深度学习描述正确的是

人工智能包含机器学习，机器学习包含深度学习

8.一般来说，逻辑回归常常用来预测二分类问题

9.支持向量机(SVM)中的支持向量是指位于间隔边界上的数据点

10.SVM的核心思想是指最大化不同类别之间的间隔

11.无监督学习中的聚类算法试图将数据点分成不同的组，这些组是基于数据点之间的相似性定义的

12.随机森林算法是通过多个决策树的预测结果结合提高模型的准确性和稳定性

13.机器学习的主要目的是让计算机从数据中生成模型，以实现某些任务或预测未知数据

14.SVM在解决非线性问题上十分出色

15.神经网络中的隐藏层主要负责识别图像中的部分特征

16.在计算机视觉领域中，深度学习被用在图像分类任务中

17深度学习模型的训练通常需要大量标记数据

18.数据压缩不是机器学习的主要任务

19.在深度学习中，数据集常常被划分为训练集、测试集、验证集。

20.测试集用于模型的最终性能评估

21.训练集作用：调整模型参数

22.验证集作用：选择最优参数和结构

23.在深度学习中，数据可以分为结构化数据和非结构化数据

24.在深度学习中，数据集的每个样本由输入数据和标签组成

25.数据标注的主要目的是：将原始数据转化为结构化数据，便于机器学习模型和理解

26.手动标注数据标注方法最适合处理复杂或特定领域的数据

27.数据标记的主要目的是：为每个数据样本分配标签，以帮助模型学习和预测

28.在深度学习中，数据集是由多个数据样本的集合组成的。

29.数据增强的作用是：增加数据集的多样性

30.在深度学习中，选择和设计合适的模型时，模型的颜色因素不是主要考虑的因素

31.过拟合：模型在训练集上表现很好，但在验证集和测试集上表现很差

32.在处理图像时，卷积神经网络最适合用于图像分类、目标检测和图像分割

33.欠拟合：模型在训练集和验证集上表现都不好

34.模型评估的主要目的是：确保模型在实际应用中的表现

35.在模型训练过程中，为了防止过拟合，需要采取：使用验证集进行评估：使用更多的训练数据

36.在神经网络中，神经元之间的连接权重决定了神经元将如何改变及传递给哪些神经元

37.KNN算法的核心思想：找到特征空间当中最邻近的K个样本，并根据这些样本的类别来决定新数据的类别

38.在K均值算法中，初始聚类中心的选择对最终结果的影响较大

39.K均值算法的目标函数是最小化簇内距离

40.K均值算法迭代过程中的两个主要步骤是计算距离和更新中心

41.K均值算法对远离簇中心类型的异常值特别敏感

42.K均值算法适用于球形或任意凸形的数据分布

43.K均值算法属于无监督学习的机器学习方法

44.在数据预处理阶段，特征选择不是常见的数据清洗操作

45.在数据预处理中，离群点检测可以用来处理异常值

46.Ordinal编码适用于将有序类别变量转换为数值表示

47.在数据预处理中，降维可以用来处理高维数据

48.在箱线图中，下四分位数指数据集中25%位置的数值

49.箱线图中的上须通常表示上四分位加上1.5倍的四分位距

50.在箱线图中中位数位于数据集中50%位置的数值

51.如果一个箱线图特别短，这可能意味着数据集的最小值接近下四分位数

52.在绘制热力相关系数图时，如果两个变量之间的相关性非常强，那么它们对应的单元格颜色会较浅

53.热力相关系数图中，通常用红色表示正相关

54.热…，如果两个变量之间的相关系数为0，那么它们对应的单元格颜色会较浅

55.箱线图适合反映分布情况和离散情况

56.热…用于展示多个变量之间的相关系数

57.回归问题属于有监督学习

58.随机森林回归是一种集成学习方法类型的模型

59.MAE平均绝对误差：预测值与真实值差的绝对值的平均值，主要用于衡量模型的预测误差大小

60.均方根误差（RMSE）与均方误差（MSE）相比，有什么不同：RMSE是MSE的平方根

61.传统人工神经网络（ANN）的主要缺点是：无法有效处理图像数据的空间信息

62.卷积神经网络（CNN）的卷积层负责提取图像的局部特征。

63.主要处理结构化数据不是卷积神经网络的特点

64.CNN能够自动学习图像的层次化信息

65.CNN通过权值共享和局部连接减少了参数数量，因此计算效率较高

66.CNN最适合图像识别任务

67.CNN最重要的一个特点是：考虑了特征之间的相关性

68.池化操作的主要目的是：减少卷积特征的空间大小，降低计算量

69.关于最大池化：保留检测区域的最显著特征

70.循环神经网络（RNN）主要特点：具有内部记忆，可以处理时间依赖性数据

71.在每个时间步，RNN如何更新其隐藏状态

基于当前输入和未来的时间步输入

72.RNN：由于梯度消失和爆炸问题，RNN的训练过程可能非常困难，需要仔细地参数调整

73.生成对抗神经网络（GAN）中的生成器的主要任务是：学习如何从随机噪声中生成“以假乱真”的数据。生成器的目的是：生成越来越像真实样本的数据。

74.GAN中的判别器的主要任务是：区分生成器生成的假样本和真实数据

75.随着GAN的训练进行，生成器和判别器的能力会发生什么变化：

生成器生成的数据越来越真实，判别器的能力也会逐渐提高