



# 职场办公AI效率革命

## 基于生成式AI的自动化效能体系构建

主讲人：朱一亨

南京农业大学 人工智能学院

2025年04月12日

# 目录

01 PPT制作

02 会议纪要

03 数据处理

04 总结展望

PART

01

PPT制作



# AI工具简介

01

## DeepSeek

DeepSeek是一款专注于学术内容生成的先进AI工具，具备强大的文献处理、数据分析等功能，能够根据用户输入的主题，快速生成高质量的学术观点和研究案例，为学术研究提供有力支持。

02

## Kimi

Kimi是一款智能PPT设计工具，提供丰富多样的模板和智能版式推荐，通过对话式交互，帮助用户快速构建PPT框架，显著提升设计效率，让演示文稿制作更加轻松高效。

# 工作模式一：内容生成与设计同步

## ■ 操作流程

使用DeepSeek生成学术内容，直接导入Kimi进行PPT设计，实现内容与设计的无缝对接。  
Kimi根据内容自动推荐版式和模板，快速构建PPT框架，提升制作效率。

## ■ 优势与适用场景

提升制作效率，适合时间紧迫的学术演示准备，快速生成高质量PPT。  
适用于科研报告、学术讲座等场景，满足多样化需求。

## ■ 实操案例

展示一个完整的从内容生成到PPT设计的案例，如“本科毕业论文答辩”



# 制作本科毕业论文答辩PPT

- 将毕业论文题目输入到 deepseek 中（输入内容要求尽可能详细）



我是南京农业大学人工智能学院的一名大四毕业生，专业是计算机科学与技术。我的毕业论文题目是“基于深度学习的蛋白质功能预测方法研究与应用”，请帮我制作一份毕业论文答辩PPT大纲，不需要数学公式和流程图，格式为markdown。



markdown

复制

## # 毕业论文答辩PPT大纲

\*\*题目：基于深度学习的蛋白质功能预测方法研究与应用\*\*  
\*\*姓名：XXX\*\*  
\*\*学号：XXX\*\*  
\*\*学院：人工智能学院\*\*  
\*\*专业：计算机科学与技术\*\*  
\*\*指导教师：XXX\*\*

### ## 1. 研究背景与意义

- \*\*蛋白质功能预测的重要性\*\*
  - 生物学、药物设计等领域的关键问题
  - 传统实验方法成本高、耗时长
- \*\*深度学习的优势\*\*
  - 自动提取特征、处理高维数据
  - 在生物信息学中的成功应用案例



# 对deepseek的输出结果进行修改和优化

基于深度学习的蛋白质功能预测研究

markdown 复制

- 数据收集与预处理 (如UniProt数据库)
- 深度学习模型设计与优化 (如CNN、Transformer等)
- 实验结果分析与对比

## ## 3. 研究方法与技术

- **数据预处理**
  - 蛋白质序列编码方法 (如氨基酸嵌入、One-Hot编码)
  - 数据集划分与增强技术
- **模型架构**
  - 核心网络结构选择 (如ResNet、LSTM)
  - 多任务学习或注意力机制的应用 (简要说明)

## ## 4. 实验结果与分析

- **实验设置**
  - 评估指标 (如准确率、F1-score)
  - 对比基线模型 (如BLAST、SVM)
- **结果展示**
  - 主实验结果的表格/图表 (如模型性能对比)
  - 关键发现与局限性分析

## ## 5. 应用与展望

- **实际应用场景**

第三节“研究方法与技术”中，增加大语言模型和transformer的内容

深度思考 (R1) 联网搜索



基于深度学习的蛋白质功能预测研究

markdown 复制

- 构建高效准确的蛋白质功能预测模型
- 探索深度学习在生物信息学中的应用潜力
- **研究内容**
  - 数据收集与预处理 (如UniProt数据库)
  - 深度学习模型设计与优化 (如CNN、Transformer等)
  - 实验结果分析与对比

## ## 3. 研究方法与技术

- **数据预处理**
  - 蛋白质序列编码方法 (如氨基酸嵌入、One-Hot编码)
  - 数据集划分与增强技术
- **模型架构**
  - **传统深度学习模型**
    - CNN、LSTM在蛋白质序列分析中的应用
  - **Transformer与大语言模型**
    - Transformer的自注意力机制在长序列建模中的优势
    - 预训练大语言模型 (如ProtBERT、ESM) 的迁移学习策略
    - 针对蛋白质功能的微调方法设计
  - **多模态融合 (可选)**
    - 结合结构预测模型 (如AlphaFold) 的输出特征

## ## 4. 实验结果与分析

- **实验设置**

## ■ 将 deepseek 生成的内容输入到kimi中

The screenshot shows the Kimi+ application interface. On the left is a vertical sidebar with several icons: a 'K' logo, a chat bubble, a document, a cluster of nodes (highlighted with a red arrow), a blue speech bubble with a red notification badge, a 'P' icon (highlighted with a red arrow), a person icon, a mobile phone icon, and a monitor icon. The main content area features the 'KIMI+' logo at the top, a greeting message, and a navigation bar with tabs: '我的关注' (selected), '官方推荐', '办公提效', '辅助写作', '社交娱乐', and '生活实用'. Below the navigation bar, there are three sections: '我的关注' containing a 'Kimi 探索版' card; '官方推荐' containing three cards: 'Kimi 探索版', '长文生成器', and '爆款网文生成器'. A red box highlights the 'PPT 助手' card in the '官方推荐' section, which is described as 'Kimi + AiPPT = 一键生成PPT' from '来自AiPPT.cn'.

**KIMI+**

👋 大家好，这里是 Kimi 的新朋友们  
和 Kimi 聊天时，随时 @ Kimi+，召唤你的私人助理

我的关注    官方推荐    办公提效    辅助写作    社交娱乐    生活实用

### 我的关注

**Kimi 探索版**  
问题太复杂？用探索版搜搜看  
来自Kimi

### 官方推荐

**Kimi 探索版**  
问题太复杂？用探索版搜搜看  
来自Kimi

**长文生成器**  
你说一句话，我写万字长文  
来自Kimi

**爆款网文生成器**  
重生之——我在Kimi写网文  
来自犄角啊犄角

**PPT 助手**  
Kimi + AiPPT = 一键生成PPT  
来自AiPPT.cn



PPT 无处不在, Kimi+为你分忧, 帮你制作详尽且有说服力的 PPT 文稿, 转达信息, 说服听众

以“中国传统文化艺术的魅力”为主题, 生成一份PPT >

我是个游戏行业从业者, 帮生成一份王者荣耀VS英雄联盟的PPT >

帮我制作一份竞选学生会主席的PPT >

### # 毕业论文答辩PPT大纲

**\*\*题目: 基于深度学习的蛋白质功能预测方法研究与应用\*\***

**\*\*姓名: XXX\*\***

**\*\*学号: XXX\*\***

**\*\*学院: 人工智能学院\*\***

**\*\*专业: 计算机科学与技术\*\***

**\*\*指导教师: XXX\*\***

联网搜索



## ## 6. 总结与致谢

- \*\*研究成果总结\*\*
  - 实现了XX性能提升
  - 验证了深度学习在该领域的可行性
- \*\*致谢\*\*
  - 导师与实验室的支持
  - 同行研究者的工作启发

# 基于深度学习的蛋白质功能预测方法研究与应用

## 1. 研究背景与意义

### 1.1 蛋白质功能预测的重要性

#### 1.1.1 生物医学领域的关键

- 蛋白质功能是生物医学研究的核心，其准确预测有助于理解生命活动机制，为疾病诊断提供理论基础。
- 在药物研发中，明确蛋白质功能可精准定位药物靶点，加速新药研发进程，降低研发成本与风险。

### 1.2 深度学习的优势

## 6.1.1 性能提升与可行性验证

- 实现了相比传统方法XX%的性能提升，证明了深度学习在蛋白质功能预测领域的巨大潜力。
- 通过大量实验验证了模型的准确性和稳定性，为后续研究奠定了坚实基础。

## 6.2 致谢

### 6.2.1 导师与实验室支持

- 感谢导师在研究过程中给予的悉心指导和宝贵建议，帮助克服重重困难。
- 感谢实验室提供的科研环境和资源支持，为研究顺利开展提供了有力保障。

### 6.2.2 同行研究者启发

- 感谢同行研究者的工作启发，他们的研究成果为本研究提供了重要的参考和借鉴。

复制 再试一次 分享

一键生成PPT >

实验结果中如何展示模型性能对比? >

多任务学习在模型中具体是如何应用的? >

未来改进方向中提到的多模态数据融合具体指哪些数据? >



热门推荐

历史模板

选择一套模板，开始创建 PPT

生成PPT



模板场景: 全部场景 总结汇报 教育培训 医学医疗 营销推广 商业计划 **高校专区** 企业介绍 党政宣传 自我介绍 >

设计风格: **全部风格** 简约 商务 科技 插画 政务 弥散 潮流 国风 立体 复古

主题颜色:

## 工作模式二：先设计后填充内容

### ■ 操作流程

使用Kimi构建PPT框架，确定版式和风格，打造整体视觉效果。

使用DeepSeek生成内容，填充到Kimi设计的框架中，确保内容与设计完美融合。

### ■ 优势与适用场景

适合对PPT设计有明确要求的场景，如求职面试，便于控制整体风格和逻辑。

通过先设计后填充内容，确保PPT结构清晰、内容完整。

### ■ 实操案例

以“本科毕业生求职面试”为主题，展示从设计到内容填充的过程，突出设计优势和内容适配性。



# 制作本科生求职面试PPT

## ■ 将自身情况、求职公司和岗位输入到kimi中

我是南京农业大学人工智能学院的一名大四毕业生，专业是计算机科学与技术。我准备面试阿里公司的算法工程师岗位，请帮我制作一份求职面试PPT。



### 阿里算法工程师求职面试PPT大纲

#### 1. 自我介绍

##### 1.1 个人简介

###### 1.1.1 教育背景

- 南京农业大学人工智能学院计算机科学与技术专业，系统学习数据结构、算法设计等课程，具备扎实的编程基础。
- 参加过多次编程竞赛，获得校级一等奖，锻炼了代码实现能力和问题解决能力。
- 熟悉Linux操作系统，掌握Shell脚本编写，能够高效完成服务器端任务自动化。

###### 1.1.2 技能清单

- 熟练掌握Python、Java编程语言，熟悉Linux操作系统，具备良好的代码风格和调试能力。
- 掌握深度学习框架如TensorFlow、PyTorch，能够独立完成模型搭建与训练。
- 熟悉SQL语言，能够高效进行数据库操作，优化查询性能，处理复杂数据关系。



# ■ DeepSeek生成内容，填充到Kimi设计的框架中。

我是南京农业大学人工智能学院的一名大四毕业生，专业是计算机科学与技术。我准备面试阿里公司的算法工程师岗位，在面试的PPT中，关于机器学习和深度学习这部分，我应该重点介绍哪些内容？



在面试阿里算法工程师岗位时，你的PPT需要突出**技术深度**和**与岗位的匹配度**。以下是针对**机器学习 & 深度学习**部分的重点内容建议，结合阿里业务场景（如电商、云计算、生物计算等）设计逻辑：

## PPT核心结构建议

### 1. 机器学习基础能力（体现扎实基础）

- **重点内容：**
  - **特征工程能力**
    - 举例：如何处理高维稀疏特征（如电商用户行为数据）
    - 技术关键词：特征选择（IV值、卡方检验）、Embedding技术
  - **模型调优经验**
    - 举例：如何解决过拟合（正则化、早停、交叉验证）
    - 技术关键词：GridSearch/RandomSearch、贝叶斯优化
  - **业务落地案例**（如有）



### 2. 深度学习核心技术（突出阿里相关技术栈）

- **重点内容：**
  - **序列建模能力**（阿里场景：用户行为序列、NLP）
    - 举例：LSTM/Transformer在时序预测/NLP任务中的应用
    - 技术关键词：Self-Attention、位置编码、Mask机制
  - **大模型相关经验**（加分项）
    - 举例：蛋白质功能预测中微调BERT/ESM的经验 → 关联阿里云生物计算场景
    - 技术关键词：Prompt Tuning、LoRA、模型蒸馏
  - **分布式训练经验**（如有）
    - 举例：多GPU训练技巧（梯度累积、混合精度）

### 3. 与阿里业务的结合点（展示业务思维）

- **电商场景：**
  - 用户画像建模（聚类/图神经网络）
  - 推荐系统（协同过滤、DeepFM）
- **云计算场景：**
  - 模型部署优化（ONNX、TensorRT）
  - 自动机器学习（AutoML）

## 工作模式三：内容优化与设计调整并行

### ■ 操作流程

使用DeepSeek生成初稿内容，导入Kimi进行初步设计，快速搭建PPT基础框架。  
根据Kimi的反馈，优化内容，调整设计细节，通过多次迭代提升PPT整体质量。

### ■ 优势与适用场景

适合对内容和设计都有较高要求的场景，如学术汇报，确保PPT在内容和形式上都达到高水平。  
通过内容优化与设计调整并行，实现高质量演示文稿的制作。

### ■ 实操案例

以“本科毕业论文答辩”为主题，展示内容优化与设计调整的过程，详细讲解优化技巧和调整方法。

# ■ 将deepseek生成的内容输入到kimi中，并进行修改和优化

## 3.2.2 多任务学习与注意力机制

- 引入多任务学习，同时预测蛋白质的多个功能类别，提高模型的综合性能，增强其在实际应用中的适应性。
- 应用注意力机制，使模型能够自动聚焦于蛋白质序列中的关键片段，提高预测的准确性和可靠性。

## 3.2.3 模型融合与集成

- 采用模型融合技术，将多个不同结构的深度学习模型进行集成，进一步提高模型的预测性能和泛化能力。
- 探索不同模型融合策略，如加权平均、投票机制等，找到最优的模型融合方案。

## 4. 实验结果与分析

### 4.1 实验设置

#### 4.1.1 评估指标

- 选择准确率、召回率、F1-score等常用指标评估模型性能，准确率反映模型预测正确的比例，召回率关注模型对正样本的识别能力，F1-score综合考虑准确率和召回率，提供更全面的性能评价。
- 采用ROC曲线和AUC值进一步评估模型的分性能，AUC值越大，模型的分性能越好。

#### 4.1.2 对比基线模型

- 将所构建的深度学习模型与传统方法（如BLAST、SVM）进行对比，BLAST基于序列相似性进行功能预测，SVM是一种经典的机器学习方法，通过对比展示深度学习模型的优势。

3.2节中，去掉多任务学习与注意力机制，增加大语言模型和transformer；4.1.2节，增加更多对比基线模型。

联网搜索



## 3.2 模型架构

### 3.2.1 核心网络结构选择

- 选择ResNet作为核心网络结构，其残差连接机制可有效解决深层网络训练中的梯度消失问题，提高模型的训练效果。
- Transformer架构凭借自注意力机制能够捕捉蛋白质序列中的长距离依赖关系，为模型提供更丰富的序列特征。

### 3.2.2 大语言模型的应用

- 大语言模型在自然语言处理领域取得了显著成果，其强大的语言理解能力为蛋白质功能预测提供了新的思路。
- 将大语言模型应用于蛋白质功能预测，通过预训练和微调，能够更好地捕捉蛋白质序列的语义信息。

### 3.2.3 模型融合与集成

- 采用模型融合技术，将多个不同结构的深度学习模型进行集成，进一步提高模型的预测性能和泛化能力。
- 探索不同模型融合策略，如加权平均、投票机制等，找到最优的模型融合方案。

### 4.1.2 对比基线模型

- 将所构建的深度学习模型与传统方法（如BLAST、SVM）进行对比，BLAST基于序列相似性进行功能预测，SVM是一种经典的机器学习方法，通过对比展示深度学习模型的优势。
- 选择更多不同的基线模型进行对比，如随机森林、朴素贝叶斯等，确保对比结果的全面性和可靠性。

PART

02

会议纪要



## 会议纪要整理流程

会议视频/录音  $\xrightarrow{\text{AI转写工具}}$  会议文档  $\xrightarrow{\text{Deepseek}}$  结构化的会议纪要

### 常用的AI转写工具：

- 通义听悟
- 讯飞听见
- 腾讯云语音识别
- 钉钉/飞书/Zoom 自带录音转文字

案例分析：博士论文答辩会议纪要整理



# 博士答辩会议纪要

## ■ 将会议视频/音频上传到通义听悟，转换成会议文字

The screenshot displays the Tongyi Tingwu (通义听悟) web interface. The main navigation menu on the left includes: 首页 (Home), 我的记录 (My Records) with a notification badge, 我的收藏 (My Favorites), 我的分享 (My Shares), 发现 (Discover), 回收站 (Recycle Bin), 通义浏览器插件 (Tongyi Browser Plugin), 低代码应用模板 (Low-code Application Templates), and 帮助与反馈 (Help & Feedback). The main content area features a search bar, a navigation bar with '企业API 全面降价, 支持自定义prompt', and a central heading '听悟一下, 随心“读”播客'. Three primary action cards are visible: '开启实时记录' (Start Real-time Recording), '上传音视频' (Upload Audio/Video) which is highlighted with a red arrow, and '播客链接转写' (Podcast Link Transcription). Below these, a '最近' (Recent) section shows three meeting records with titles like 'meeting\_02' and '【示例】上课用通义听悟, 学习效率UPUP', along with tags for '课程', '效率', '音视频', and '转文字'.

## 上传本地音视频文件



点击 / 拖拽 本地音视频文件到这里

- 单个文件最长6小时，单次最多可上传50个。
- 视频支持：mp4/wmv/m4v/flv/rmvb/dat/mov/mkv/webm/avi/mpeg/3gp/ogg，单个最大6G；
- 音频支持：mp3/wav/m4a/wma/aac/ogg/amr/flac/aiff，单个最大500M。

音视频语言 ?

中文

英语

日语

粤语

中英文自由说

翻译

不翻译

区分发言人

暂不体验

单人演讲

2人对话

多人讨论

开始转写

取消

Q 搜索...

低代码应用模板，开箱即用可私有化部署



上传中(1)

## 听悟一下，轻松学习

**会议神器**

### 开启实时记录

实时语音转文字  
同步翻译，智能总结要点

**网课必备**

### 上传音视频

音视频转文字  
区分发言人，一键导出

**AI看播客**

### 播客链接转写

输入 RSS 订阅链接  
无需下载，智能提炼总结

全选

meeting\_02 ×

位置：默认文件夹 已上传(53%)

最近

**meeting\_02**

基因本体论 机器学习 度量学习

01:14:54 今天 17:23

**【示例】上课用通义听悟，学习效率UPUP**

课程 效率 音视频 转文字

01:53 今天 17:22

**【示例】开会用通义听悟，高效又省心**

会议 纪要 实时记录 翻译 议程

01:47 今天 17:22

发言人 1 00:00

教授以及天津大学的杜浦峰教授，来作为朱一峰同学的博士学位论文答辩委员会，组成这个答辩委员会。我们就有请沈教授来作为一下这个主席，下面我简单介绍一下注意课程的一个基本情况。朱毅恒，它实际上是 2015 年 6 月在南京工程学院获得工学学士学位。同年 9 月，就进到我们南理工计算机学院来攻读硕士的，后来，就是转为博士，19 年 12 月到 21 年，这两 12 月，两年时间在密歇根大学就张扬教授的一篇实验室来做访问。在攻读博士学位期间，他主要是做蛋白质功能预测的。在这个 GPB 以及 BIB 了 CB 上面发表了几篇相关的文章。今天的他的这个小论文了，大论文已经达到我们学校的这样一个毕业要求，大论文的外审也已经正常通过了。所以今天请各位专家再把把关，给他多提宝贵意见，一横再做相应的修改，最后完成这样一个博士学位论文。好，下面我们就有请沈教授来组织这样一场答辩，好吧。

发言人 2 01:25

谢谢于老师还有一个就是所有专家的一个露脸截屏，需要一个。

发言人 1 01:32

好的，那麻烦各位专家能露下脸。好。

发言人 2 01:45

好。

发言人 1 01:46

可以了吧？

发言人 2 01:48

可以了。

发言人 1 01:49

好的，那陈教授下面就委托您了。

发言人 4 01:53

好的，谢谢于教授。下面就请朱一恒同学把博士阶段的工作做一个简单的一个汇报，大概在 30 到 40 分钟之间好吧，是吧？我不知道南理工的要求是什么样子的。



已保存



保存



导出



得时长

分享



收藏



更多

### 全文概要

在一场精心组织的答辩会上，一个由经验丰富的专家组成的委员会对一名研究生的博士学位论文进行了深入评审。该论文聚焦于利用机器学习与基因本体论进行蛋白质功能预测的研究，彰显了研究者在生物信息学领域的深厚学术背景与创新思维。答辩过程中，委员们不仅详细询问了研究的每一个细节，还就如何进一步提升方法的创新性、增强模型的可解释性以及规划后续研究方向提出了建设性的意见。基于论文的严谨性与研究者的出色表现，答辩委员会以高度一致的意见，无记名投票方式，决定通过论文答辩，推荐授予工学博士学位，这不仅是对研究者学术成果的认可，也为未来的研究指明了方向。

收起

### 章节速览

发言总结

问答回顾

00:00

#### 朱一恒博士学位论文答辩会议

本次对话围绕朱一恒同学的博士学位论文答辩展开，介绍了其学术背景，包括在南京工程学院获得学士学位，随后在南京理工大学攻读硕士并转为博士，以及在密歇根大学的访问经历。朱一恒的研究主要集中在蛋白质功能预测领域，已发表多篇相关论文，并满足学校毕业要求。会议邀请了多位专家组成答辩委员会，由沈教授担任主席，对朱一恒的博士学位论文进行评审和指导，旨在帮助其进一步完善论文，顺利毕业。朱一恒将在接下来的30到40分钟内汇报其博士阶段的研究工作。

02:35

#### 基于机器学习与基因本体论的蛋白质功能预测研究

该研究聚焦于基于机器学习与基因本体论的蛋白质功能预测，旨在解决蛋白质功能注释的低效问题。蛋白质在生命过程中扮演关键角色，准确识别其功能对解释生命活动、疾病机理和药物设计至关重要。基因本体论作为功能注释的语义词汇标准，涵盖了分子功能、生物过程和细胞组件三个分支。然而，传统的生物实验方法存在周期长、成本高的局限，导致功能注



00:00

01:14:54

倍速



## ■ 将会议文字文档上传到deepseek上，生成结构化会议纪要



meeting\_02\_原文.docx

DOCX 31.65KB

将对话内容转换成结构化的会议纪要，并列专家的提问及答辩人的回答。

### 结构化会议纪要：朱一恒博士学位论文答辩

时间：2025年04月06日

地点：线上会议

答辩人：朱一恒（南京理工大学计算机学院）

论文题目：基于机器学习与基因本体论的蛋白质功能预测研究

#### 一、答辩内容摘要

##### 1. 研究背景与意义

- 蛋白质功能预测是生物信息学核心问题，传统实验方法成本高、效率低。
- 基因本体论（GO）是功能注释标准，包含分子功能、生物过程、细胞组件三个分支。
- 深度学习可解决数据匮乏问题，提升预测效率。

Q1. 注意力机制是否对问题敏感？

回答：目前直接使用 ESM 的预训练框架，未来可设计领域特异性注意力机制。

#### 4. 其他专家提问

Q1（刘老师）：基因表达数据来源？

回答：来自公共数据库（如 COXPRESdb），部分数据需针对物种单独处理。

Q2（杜老师）：线性集成中权重如何设定？

回答：权重通过训练数据拟合（如神经网络层权重约 0.6），但提升效果有限（<1%）。

#### 三、答辩决议

- 论文创新性显著，成果具有理论价值与应用潜力。
- 答辩表现优秀，回答问题准确，展现扎实的专业能力。
- 全票通过答辩，建议授予工学博士学位。

#### 四、后续改进建议

- 论文优化：统一术语格式，简化方法命名。
- 技术扩展：探索结构信息融合、多模态语言模型。
- 业务结合：与阿里云生物计算平台对接，验证方法落地性。



# 对会议纪要进行修改和优化

答辩人名字错了，将“朱一恒”改为“朱一亨”

实验结果部分内容偏少，增加一些



## 结构化会议纪要：朱一亨博士学位论文答辩

时间：2025年04月06日

地点：线上会议

答辩人：朱一亨（南京理工大学计算机学院）

论文题目：基于机器学习与基因本体论的蛋白质功能预测研究

### 一、答辩内容摘要

#### 1. 研究背景与意义

- 蛋白质功能预测是生物信息学核心问题，传统实验方法成本高、效率低。
- 基因本体论（GO）是功能注释标准，包含分子功能、生物过程、细胞组件三个分支。
- 深度学习可解决数据匮乏问题，提升预测效率。

#### 2. 主要研究工作

- 基于基因视角：提出 **TripletGO** 方法，融合基因表达、序列等多源数据，通过三元组神经网络度量表



## 三、实验结果与分析（补充版）

### 1. TripletGO方法性能

- 数据集**：在自构建的基因数据集和CAFA3竞赛数据集上测试
- 对比方法**：与5种无监督基因表达相似性度量方法比较
- 关键结果**：
  - 在编码基因数据集上，F1-score达到0.78，比最优基线提升12% ( $p < 0.01$ )
  - 对稀有GO术语（出现频率 $< 30$ ）的预测AUC提升15-20%
  - 四类生物信息源贡献度分析显示：基因表达数据（32%）>蛋白质序列（28%）>基因序列（25%）>GO术语频率（15%）

### 2. ATGO方法表现

- 预训练效果**：
  - ESM Transformer在千万级未标注序列上预训练后，下游任务准确率提升23%
  - 消融实验表明：使用最后三层特征融合比单层特征AUC高8%



PART

03

数据处理



# 数据生成

帮我把以上数据转换为.csv格式

## ■ 表格数据（南京市天气）

南京市近七天天气情况，以表格形式展现。

以下是南京市未来7天天气预报（2025年4月7日-4月13日，模拟数据仅供参考，实际天气请以权威气象台为准）：

日期	白天天气	夜间天气	气温范围 (°C)	风向风力	降水概率	空气质量指数
4月7日	多云	阴	12 ~ 20	东南风3级	10%	良 (75)
4月8日	小雨	中雨	14 ~ 18	东风4级	80%	优 (50)
4月9日	阴转多云	晴	10 ~ 16	西北风3级	20%	良 (65)
4月10日	晴	晴	9 ~ 19	西南风2级	0%	轻度污染 (105)
4月11日	多云	雷阵雨	15 ~ 22	南风3-4级	60%	良 (85)
4月12日	大雨	小雨	13 ~ 18	东北风5级	90%	优 (40)
4月13日	阴	多云	11 ~ 17	北风3级	30%	良 (70)

温馨提示：



以下是南京市未来7天天气预报的CSV格式数据（可直接复制到文本文件并保存为.csv格式）：

```

csv
日期,白天天气,夜间天气,气温范围 (°C),风向风力,降水概率,空气质量指数
4月7日,多云,阴,"12 ~ 20",东南风3级,10%,良 (75)
4月8日,小雨,中雨,"14 ~ 18",东风4级,80%,优 (50)
4月9日,阴转多云,晴,"10 ~ 16",西北风3级,20%,良 (65)
4月10日,晴,晴,"9 ~ 19",西南风2级,0%,轻度污染 (105)
4月11日,多云,雷阵雨,"15 ~ 22",南风3-4级,60%,良 (85)
4月12日,大雨,小雨,"13 ~ 18",东北风5级,90%,优 (40)
4月13日,阴,多云,"11 ~ 17",北风3级,30%,良 (70)

```



复制

### 使用说明：

1. 将上方内容复制到文本编辑器中
2. 保存为 `南京天气_20250407-0413.csv` 或其他您喜欢的文件名
3. 确保文件扩展名为 `.csv`
4. 可用Excel/WPS/记事本等软件直接打开

### 天气

日期	白天天气	夜间天气	气温范围 (°C)	风向风力	降水概率	空气质量指数
4月7日	多云	阴	12 ~ 20	东南风3级	10%	良 (75)
4月8日	小雨	中雨	14 ~ 18	东风4级	80%	优 (50)
4月9日	阴转多云	晴	10 ~ 16	西北风3级	20%	良 (65)
4月10日	晴	晴	9 ~ 19	西南风2级	0%	轻度污染 (105)
4月11日	多云	雷阵雨	15 ~ 22	南风3-4级	60%	良 (85)
4月12日	大雨	小雨	13 ~ 18	东北风5级	90%	优 (40)
4月13日	阴	多云	11 ~ 17	北风3级	30%	良 (70)



## ■ 绘制流程图（学生成绩管理系统）

我现在要开发一款学生成绩管理系统，请帮我生成一份详细的软件开发流程图，并输出为mermaid格式



以下是为学生成绩管理系统设计的详细软件开发流程图，采用 Mermaid 格式，涵盖需求分析、设计、开发、测试到部署的全流程。可直接粘贴到支持 Mermaid 的编辑器（如 Typora、VS Code、GitHub Markdown）中生成可视化图表。

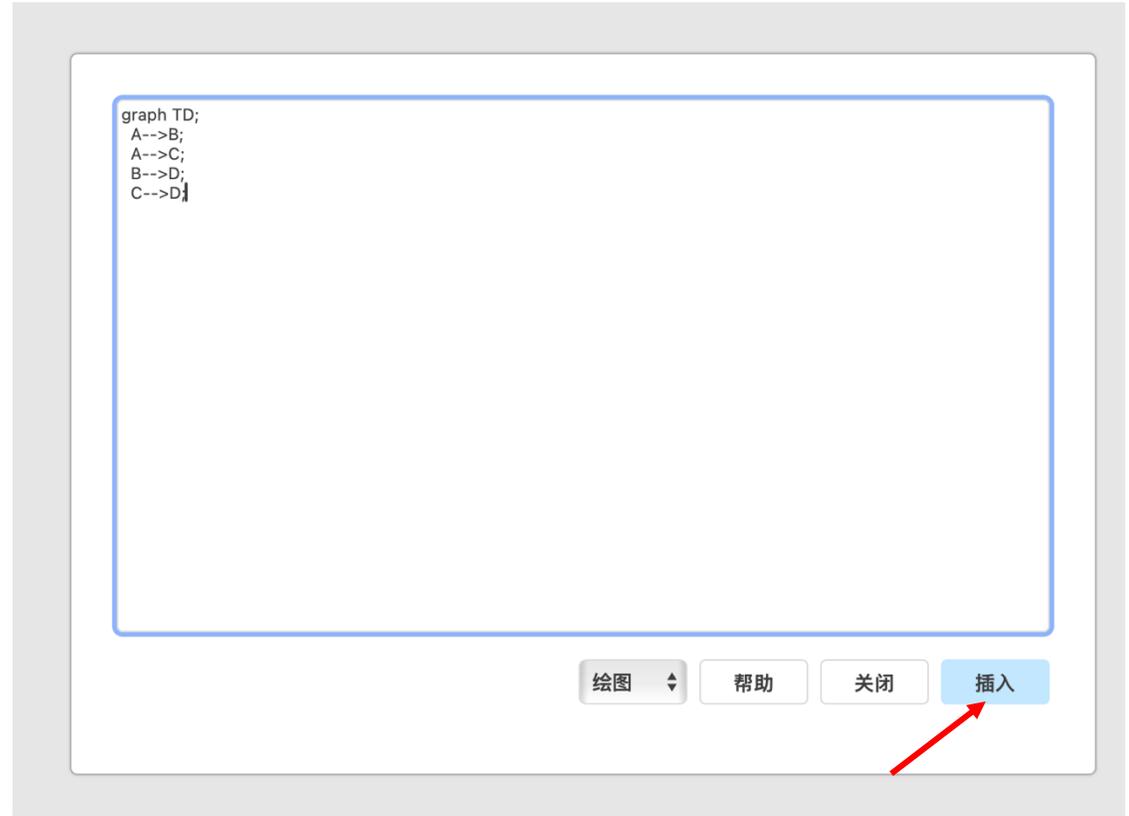
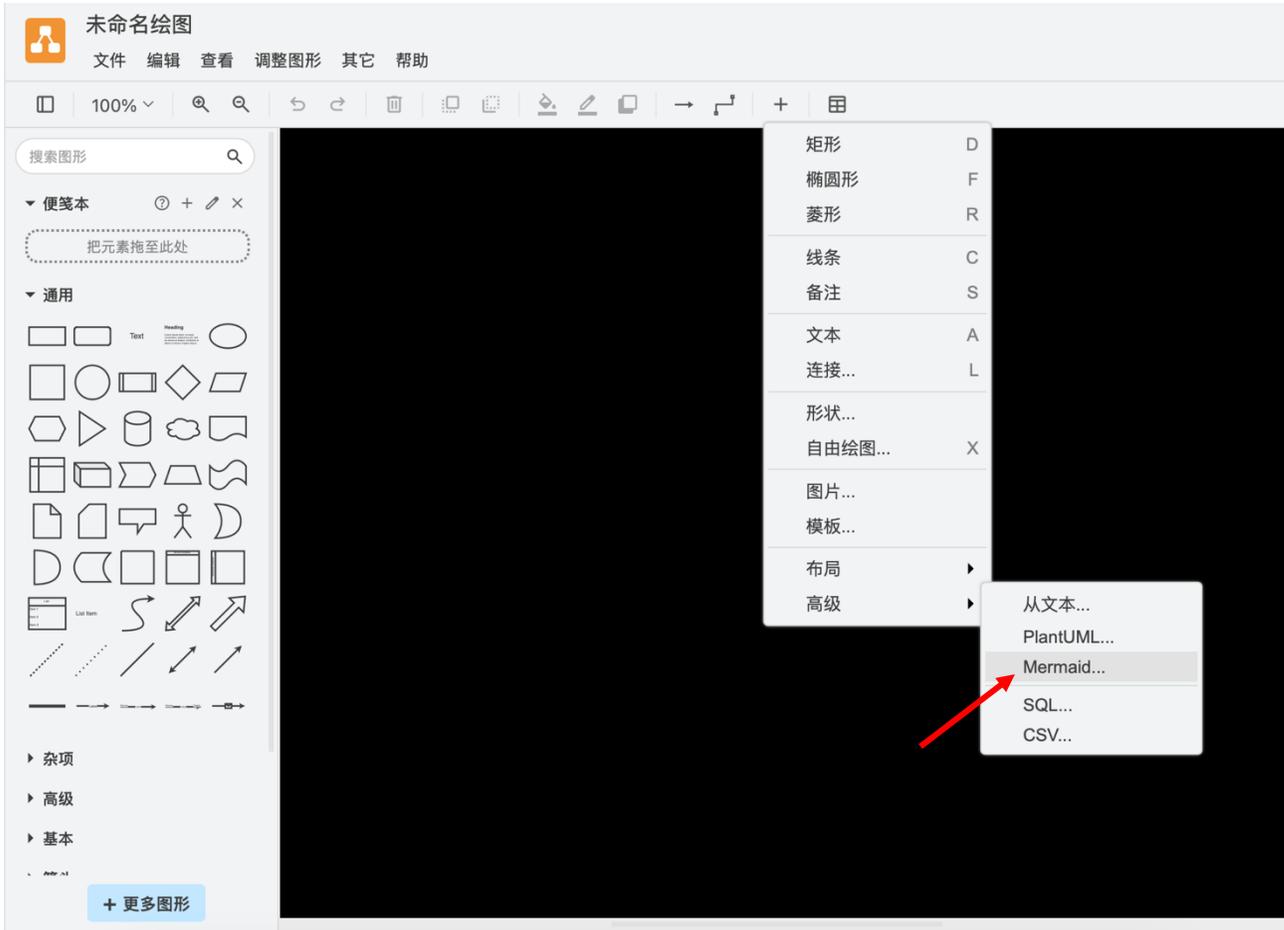
### 学生成绩管理系统开发流程图（Mermaid）

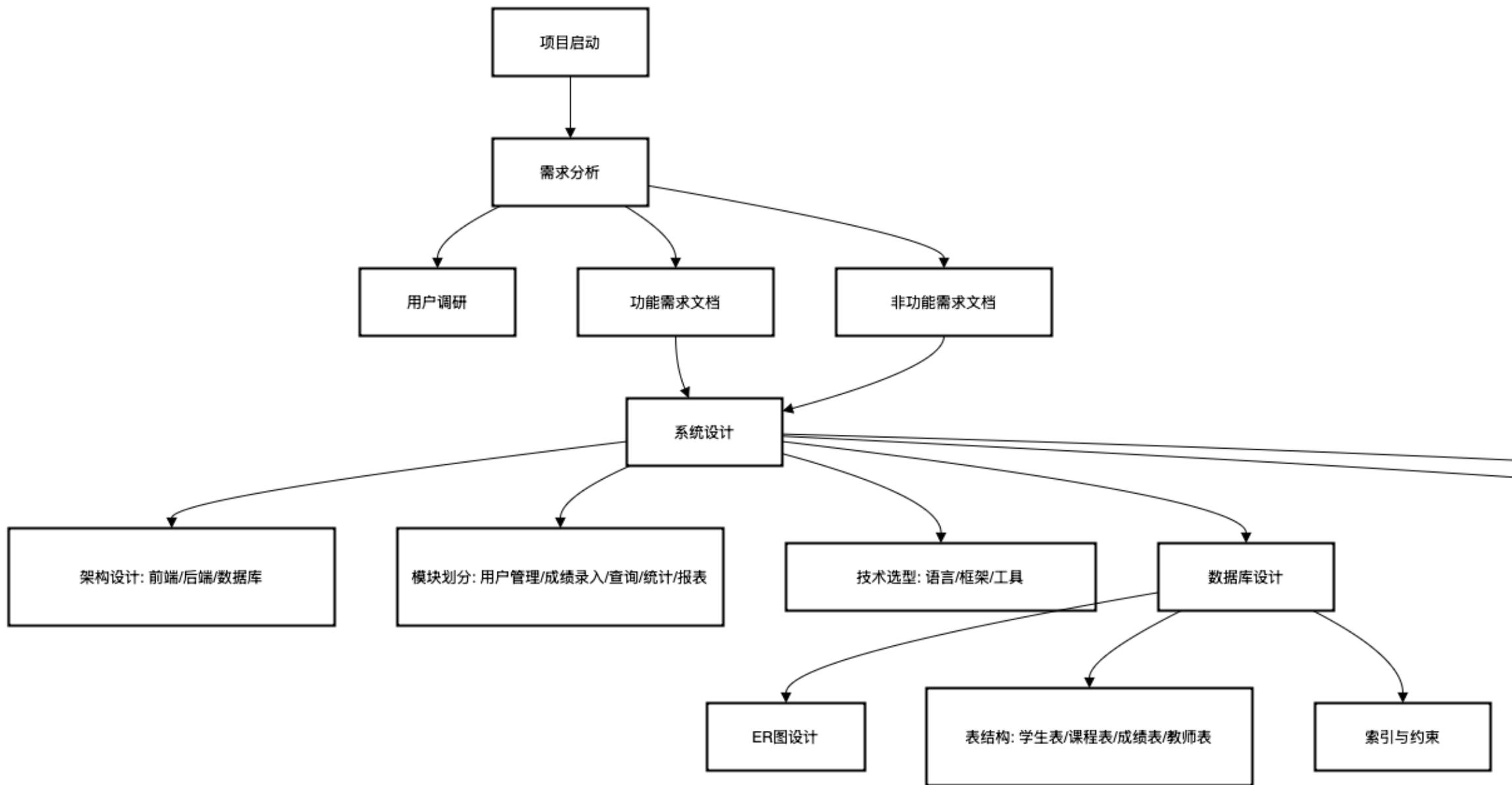
复制

```
mermaid
flowchart TD
  A[项目启动] --> B[需求分析]
  B --> B1[用户调研]
  B --> B2[功能需求文档]
  B --> B3[非功能需求文档]
  B2 --> C[系统设计]
  B3 --> C
  C --> C1[架构设计：前端/后端/数据库]
  C --> C2[模块划分：用户管理/成绩录入/查询/统计/报表]
  C --> C3[技术选型：语言/框架/工具]
  C --> D[数据库设计]
  D --> D1[ER图设计]
  D --> D2[表结构：学生表/课程表/成绩表/教师表]
  D --> D3[索引与约束]
```

```
C --> E[UI原型设计]
E --> E1[低保真原型]
E --> E2[高保真交互图]
E --> F[前端开发]
F --> F1[页面实现：Vue/React]
F --> F2[接口联调]
C --> G[后端开发]
G --> G1[API开发：Spring/Django]
G --> G2[业务逻辑：成绩计算/权限控制]
G --> G3[数据库CRUD]
F2 --> H[系统集成]
G3 --> H
H --> I[测试阶段]
I --> I1[单元测试]
I --> I2[集成测试]
I --> I3[性能测试]
I --> I4[用户验收测试]
I4 --> J[部署上线]
J --> J1[服务器配置]
J --> J2[数据库迁移]
J --> J3[监控与日志]
J --> K[维护迭代]
K --> K1[Bug修复]
K --> K2[功能扩展]
```

# ■ Diagrams.net (免费的在线图表绘制工具)







# 数据分析

## ■ 考试成绩分析

3	4 姓名	5 班级	6 考试性质	7 期末成绩			8 总成绩	9 成绩标志
				10 平时成绩	11 期末成绩	12 期末成绩		
13	张1	人工智能211	正常考试	90	40	60	60	
14	张2	人工智能211	正常考试	90	50	66	66	
15	张3	人工智能211	正常考试	90	70	78	78	
16	张4	人工智能211	正常考试	95	95	95	95	
17	张5	人工智能211	正常考试	90	90	90	90	
18	张6	人工智能211	正常考试	90	65	75	75	
19	张7	人工智能212	正常考试	90	87	88	88	
20	张8	人工智能212	正常考试	90	70	78	78	
21	张9	人工智能212	正常考试	90	60	72	72	
22	张10	人工智能212	正常考试	90	40	60	60	
23	张11	人工智能212	正常考试	90	40	60	60	
24	张12	人工智能212	正常考试	90	60	72	72	
25	张13	人工智能212	正常考试	90	81	85	85	
26	张14	人工智能212	正常考试	90	75	81	81	
27	张15	人工智能211	正常考试	90	70	78	78	
28	张16	数据211	正常考试	90	60	72	72	
29	张17	数据211	正常考试	85	60	70	70	
30	张18	数据211	正常考试	0	0	0	0	旷考
31	张19	数据212	正常考试	90	73	80	80	
32	张20	数据212	正常考试	90	70	78	78	
33	张21	数据211	正常考试	90	77	82	82	
34	张22	人工智能211	正常考试	95	40	62	62	

我是南京农业大学的一名教师，教授的课程是《生物信息学》，请帮我写一份考试成绩分析，300字左右。包含平均分，各分段人数、教学存在的不足、改进方向等。



### 《生物信息学》期末考试成绩分析报告

#### 一、成绩总体情况

本次考试共有22名学生参加（含1人旷考），经统计分析：

1. 平均分：74.8分（剔除旷考学生）
2. 分数分布：
  - 优秀（≥90分）：2人（9.1%）
  - 良好（80-89分）：3人（13.6%）
  - 中等（70-79分）：8人（36.4%）
  - 及格（60-69分）：5人（22.7%）
  - 不及格（<60分）：3人（13.6%）

#### 二、教学存在问题

1. 成绩分布不均：优秀率偏低，中等成绩学生占比较大
2. 班级差异明显：数据班级整体表现较弱，平均分低于人工智能班级
3. 学用脱节现象：部分学生（如张22）平时成绩优异但期末成绩大幅下滑

#### 三、改进措施建议

1. 优化教学设计：
  - 增加案例分析和实践环节比重
  - 针对不同班级调整教学难度
2. 加强过程管理：
  - 实施阶段性学习效果检测
  - 建立学习困难学生帮扶机制
3. 完善评价体系：
  - 细化平时成绩考核标准
  - 增加课堂互动表现评分
4. 重点关注：
  - 加强对数据班级的教学指导

# 实验数据分析

**Table 1.** The overall performance of 15 function prediction methods on all 1522 test proteins

Table1.docx  
DOCX 25.97KB

Method	F <sub>max</sub>			S <sub>min</sub>			AUPRC			Coverage <sup>f</sup>		
	MF	BP	CC	MF	BP	CC	MF	BP	CC	MF	BP	CC
Blast-KNN <sup>a, d</sup>	0.642	0.397	0.485	7.77	24.90	8.59	0.346	0.220	0.259	0.832	0.803	0.717
FunFams <sup>a, d</sup>	0.483	0.311	0.387	9.87	27.24	9.02	0.298	0.141	0.200	0.631	0.599	0.532
PPIGO <sup>a, d</sup>	0.329	0.273	0.461	11.81	26.74	8.43	0.141	0.126	0.253	0.515	0.558	0.645
DeepGOCNN <sup>b, e</sup>	0.430	0.296	0.497	11.01	26.67	9.45	0.369	0.204	0.493	1.000	1.000	1.000
TALE <sup>b, d</sup>	0.457	0.313	0.526	11.19	25.88	8.77	0.397	0.222	0.534	1.000	1.000	1.000
DeepGOZero <sup>b, d</sup>	0.677	0.396	0.540	7.53	24.86	9.46	0.674	0.319	0.521	1.000	1.000	1.000
AnnoPRO <sup>b, e</sup>	0.504	0.365	0.535	9.63	25.36	8.67	0.366	0.267	0.504	1.000	1.000	1.000
HFRGO <sup>b</sup>	0.682	0.412	0.580	7.23	23.91	8.14	0.630	0.340	0.539	1.000	1.000	1.000
ATGO <sup>c, d</sup>	0.686	0.424	0.607	7.34	23.99	7.87	0.676	0.361	0.625	1.000	1.000	1.000
DeepGO-SE <sup>c, d</sup>	0.669	0.411	0.573	7.67	24.48	9.44	0.662	0.351	0.600	1.000	1.000	1.000
PLMGO <sup>c</sup>	0.680	0.424	0.628	7.58	23.95	7.57	0.621	0.355	0.571	1.000	1.000	1.000
DeepGOPlus <sup>d</sup>	0.660	0.402	0.574	7.78	24.92	8.56	0.620	0.311	0.517	1.000	1.000	1.000
TALE+ <sup>d</sup>	0.640	0.401	0.581	8.04	24.91	8.37	0.617	0.318	0.550	1.000	1.000	1.000
ATGO+ <sup>d</sup>	0.693	0.430	0.607	7.22	23.88	8.11	0.670	0.371	0.617	1.000	1.000	1.000
MKFGO	<b>0.710</b>	<b>0.459</b>	<b>0.639</b>	<b>6.97</b>	<b>23.08</b>	<b>7.38</b>	<b>0.716</b>	<b>0.400</b>	<b>0.668</b>	1.000	1.000	1.000

这张表是不同蛋白质功能预测方法的性能比较，请结合表中的实验数据进行分析，300字。

## 蛋白质功能预测方法性能分析

通过对15种蛋白质功能预测方法的比较分析，可以得出以下结论：

### 1. 整体性能表现：

MKFGO作为复合方法，在F<sub>max</sub>（分子功能0.710、生物过程0.459、细胞组分0.639）、S<sub>min</sub>（MF 6.97、BP 23.08、CC 7.38）和AUPRC（MF 0.716、BP 0.400、CC 0.668）三个核心指标上均表现最优，显示出综合预测优势。单方法中，ATGO和PLMGO表现突出，F<sub>max</sub>分别达到0.686/0.424/0.607和0.680/0.424/0.628。

### 2. 方法类型差异：

基于PLM特征表示的深度学习（c类）整体优于传统深度学习方法（b类）和模板检测方法（a类）。例如，ATGO的F<sub>max</sub>比最佳模板方法Blast-KNN提高6.8%（MF）、6.8%（BP）和25.2%（CC），验证了预训练模型的有效性。

### 3. 预测维度特点：

所有方法在分子功能（MF）预测上表现最优，细胞组分（CC）次之，生物过程（BP）最具挑战性。最佳方法MKFGO的BP指标（F<sub>max</sub>=0.459）仍显著低于其MF指标（0.710），反映生物过程预测的复杂性。

<sup>a</sup> Template detection-based methods; <sup>b</sup> Deep learning-based methods with handcraft feature representations; <sup>c</sup> Deep learning-based methods with PLM-based feature representations; <sup>d</sup> The prediction models are re-trained on our training dataset using the author's source codes; <sup>e</sup> The prediction models are directly downloaded from author's web platforms. <sup>f</sup> Coverage is the proportion of the number of test proteins with available prediction scores divided by the total number of test proteins. Bold fonts highlight the best performer in each category.



期末考试成绩.xlsx  
XLSX 10.38KB

附件是期末考试成绩表，第一行是标题，从第二行开始是每位同学的成绩，现在需要进行以下操作：(1) 去除旷考的学生；(2) 将相同班级的学生放在一起 (3) 同一班级内的学生按照成绩降序排序。请帮忙生成 VBA代码

以下是满足您需求的VBA代码，它会：(1) 删除"旷考"学生，(2) 按班级分组，(3) 每个班级内按总成绩降序排序：

```
vba
Sub ProcessScores()
    Dim ws As Worksheet
    Dim lastRow As Long, i As Long
    Dim rng As Range, delRng As Range
    Dim classCol As Integer, scoreCol As Integer, statusCol As Integer

    ' 设置工作表
    Set ws = ThisWorkbook.Sheets("Sheet1")

    ' 确定列位置
    classCol = 2 ' B列是班级
    scoreCol = 7 ' G列是总成绩
```

复制

Excel 文件 编辑 视图 插入 格式 工具 数据 窗口 帮助

开始 插入 绘图 页面布局 公式 数据

等线 (正文) 11

B I U

续订您的订阅 若要继续使用您的 Office 应用程序，而

I26

	A	B	C	D	E	F	G	H	I
1	姓名	班级	考试性质	平时成					
2	张1	人工智能211	正常考试				60		
3	张2	人工智能211	正常考试				66		
4	张3	人工智能211	正常考试				78		
5	张4	人工智能211	正常考试				95		
6	张5	人工智能211	正常考试				90		
7	张6	人工智能211	正常考试				75		
8	张7	人工智能212	正常考试				88		
9	张8	人工智能212	正常考试				70		
10	张9	数据211	正常考试						
11	张10	人工智能212	正常考试						
12	张11	人工智能212	正常考试						
13	张12	人工智能212	正常考试		90	60	72	77	
14	张13	人工智能212	正常考试		90	81	85	85	
15	张14	人工智能212	正常考试		90	75	81	81	
16	张15	人工智能211	正常考试		90	70	78	78	
17	张16	数据211	正常考试		90	60	72	72	
18	张17	数据211	正常考试		85	60	70	70	
19	张18	数据211	正常考试		0	0	0	0	旷考
20	张19	数据212	正常考试		90	73	80	80	
21	张20	数据212	正常考试		90	70	78	78	
22	张21	数据211	正常考试		90	77	82	82	
23	张22	人工智能211	正常考试		95	40	62	62	
24	张23	数据211	正常考试		0	0	0	0	旷考
25									

拼写... 同义词库... 智能查找... 语言... 自动更正选项... 错误检查... 翻译... 检查辅助功能 显示更改 修订 合并工作簿... 保护 单变量求解... 方案... 审核 宏... Excel 加载项... 自定义键盘...

宏... 录制新宏... Visual Basic 编辑器





# 数据标记

期末考试成绩.xlsx  
XLSX 10.37KB

将班级为人工智能211的学生成绩标记为红色，给出详细的VBA代码

## VBA代码：标记人工智能211班级学生成绩为红色

以下是详细的VBA代码，用于将班级为"人工智能211"的学生成绩标记为红色：

```
vba
Sub MarkAI211StudentsRed()
    Dim ws As Worksheet
    Dim lastRow As Long
    Dim i As Long

    ' 设置要操作的工作表（这里使用活动工作表，也可以指定具体工作表）
    Set ws = ActiveSheet ' 或者使用 Set ws = ThisWorkbook.Sheets("Sheet1")

    ' 找到最后一行数据
    lastRow = ws.Cells(ws.Rows.Count, "A").End(xlUp).Row

    ' 从第5行开始遍历（根据你的数据表结构，数据从第5行开始）
```

复制



A	B	C	D	E	F	G	H	I
姓名	班级	考试性质	平时成绩	期末成绩	期末成绩	总成绩	成绩标准	
张4	人工智能211	正常考试	95	95	95	95		
张5	人工智能211	正常考试	90	90	90	90		
张3	人工智能211	正常考试	90	70	78	78		
张15	人工智能211	正常考试	90	70	78	78		
张6	人工智能211	正常考试	90	65	75	75		
张2	人工智能211	正常考试	90	50	66	66		
张22	人工智能211	正常考试	95	40	62	62		
张1	人工智能211	正常考试	90	40	60	60		
张7	人工智能212	正常考试	90	87	88	88		
张13	人工智能212	正常考试	90	81	85	85		
张14	人工智能212	正常考试	90	75	81	81		
张8	人工智能212	正常考试	90	70	78	78		
张12	人工智能212	正常考试	90	60	72	72		
张10	人工智能212	正常考试	90	40	60	60		
张11	人工智能212	正常考试	90	40	60	60		
张21	数据211	正常考试	90	77	82	82		
张16	数据211	正常考试	90	60	72	72		
张17	数据211	正常考试	85	60	70	70		
张19	数据212	正常考试	90	73	80	80		
张20	数据212	正常考试	90	70	78	78		



# 数据可视化

## ■ 期末考试成绩数据看板



期末考试成绩.xlsx  
XLSX 10.37KB

请整理并分析上传的数据，生成数据看板，以html形式输出，要求如下：

- 1、分析重点数据结论，用合适的可视化方式展现。
- 2、可视化图表要符合现代化审美，UI简洁美观，布局合理，清晰易懂。
- 3、如有必要，搭配数据解读的文字。
- 4、输出格式为标准且可正常运行的html代码，不要有任何其他的解读内容。

html

复制

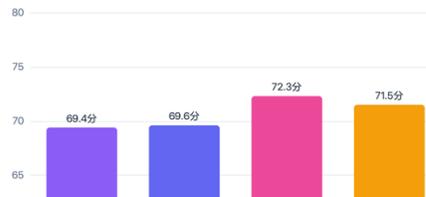
```
<!DOCTYPE html>
<html lang="zh-CN">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>期末考试成绩分析看板</title>
  <script src="https://cdn.jsdelivr.net/npm/echarts@5.4.3/dist/echarts.min.js"></script>
  <style>
    body {
      font-family: 'Arial', sans-serif;
      margin: 0;
      padding: 20px;
      background-color: #f8fafc;
    }
  </style>
</head>
</html>
```

成绩分布总览

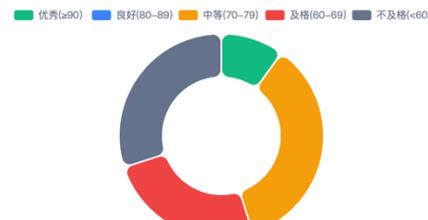


总人数: 20人 | 平均总成绩: 70.3分 | 优秀率(≥90): 10% | 不及格率(<60): 30%

班级成绩对比



成绩等级分布



## ■ 期末考试成绩小红书卡片

请整理并分析上传的数据，以html形式输出，要求如下：

- 1、符合小红书平台内容调性，制作成卡片形式，文字大小阶梯分明，清晰美观易懂。
- 2、尺寸采用小红书竖图适配布局，适合移动端阅读。
- 3、重点文字内容需要详实，不重要的可以忽略。
- 4、输出格式为标准目可正常运行的html代码，不要有任何其他的解读内容。

```
html 复制
<!DOCTYPE html>
<html lang="zh-CN">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>期末考试成绩分析报告</title>
  <style>
    body {
      margin: 0;
      padding: 0;
      background-color: #fafafa;
      font-family: "PingFang SC", "Hiragino Sans GB", "Microsoft YaHei", sans-serif;
      color: #333;
      line-height: 1.6;
    }
  </style>
</head>
</html>
```

## 成绩分析报告

### 人工智能专业期末考试成绩分析

本次共收集**20名**学生成绩数据，涵盖**4个班级**，通过可视化分析发现以下关键结论：

#### 🔍 核心发现

- 班级最高分：**95分**（张4）
- 班级最低分：**40分**（4位同学）
- 平均成绩：**70.3分**
- 不及格率：**30%**（6位同学）

#### 📊 成绩分布



成绩呈现**两极分化**趋势，优秀学生与不及格学生占比均较高

数据分析

教育研究

成绩报告

## ■ 课堂点名小程序

期末考试成绩.xlsx  
XLSX 10.37KB



帮我生成一个课堂随机点名器，网页的背景风格要大气美观。体现AI风格。点击按钮即可开始随机抽取。中间区域显示一个框框用于显示抽中的姓名，抽中的姓名显示要明显。点击开始点名按钮输入框显示出来名字滚动的效果，点击暂停按钮束的时候滚动速度变慢,增强视觉感。学生名单详见附件。



### AI 课堂随机点名器

下面是一个基于您提供的学生名单的AI风格随机点名器网页代码。这个点名器具有以下特点：

- 现代AI风格的界面设计
- 点击开始按钮时名字会滚动显示
- 点击暂停按钮时滚动速度逐渐减慢
- 抽中姓名会以明显方式显示
- 响应式设计，适应不同设备

html

复制

```
<!DOCTYPE html>  
<html lang="zh-CN">
```





# 数据推理 (鸢尾花的分类)

iris setosa



petal sepal

(花萼) (花瓣)

iris versicolor



petal sepal

iris virginica



petal sepal

```

150,4,setosa,versicolor,virginica
5.1,3.5,1.4,0.2,0
4.9,3.0,1.4,0.2,0
4.7,3.2,1.3,0.2,0
4.6,3.1,1.5,0.2,0
5.0,3.6,1.4,0.2,0
5.4,3.9,1.7,0.4,0
4.6,3.4,1.4,0.3,0
5.0,3.4,1.5,0.2,0
4.4,2.9,1.4,0.2,0
4.9,3.1,1.5,0.1,0
5.4,3.7,1.5,0.2,0
4.8,3.4,1.6,0.2,0
4.8,3.0,1.4,0.1,0
4.3,3.0,1.1,0.1,0
5.8,4.0,1.2,0.2,0
5.7,4.4,1.5,0.4,0
5.4,3.9,1.3,0.4,0
5.1,3.5,1.4,0.3,0
5.7,3.8,1.7,0.3,0
5.1,3.8,1.5,0.3,0
5.4,3.4,1.7,0.2,0
5.1,3.7,1.5,0.4,0
4.6,3.6,1.0,0.2,0

```

2.5, 2.1, 4.9, 1.5 ?

Iris也称鸢尾花卉数据集，由Fisher在1936收集整理。

数据集包含150个数据样本，分为3类：Setosa（山鸢尾），Versicolour（变色鸢尾），Virginica（维吉尼亚鸢尾），每类50个数据。

每个数据包含4个属性：花萼长度，花萼宽度，花瓣长度，花瓣宽度。

附件是UCI数据集中的iris数据集，前4列数据是特征，分别是：sepal length、sepal width、petal length和petal width；第5列是花的种类。请问，如果有一朵花的前4列特征分别是2.5, 2.1, 4.9, 1.5，它应该是什么种类？请使用机器学习算法中的支持向量机模型帮我预测一下，并直接告诉我答案。



预测结果：这朵花的种类是 **Iris-virginica**（维吉尼亚鸢尾）。

## 关键分析

### 1. 模型判断依据：

- 花瓣长度（4.9）和宽度（1.5）显著大于 **Iris-versicolor** 的典型范围（花瓣长度通常 $\leq 4.5$ ），更接近 **virginica** 的特征分布。
- 花萼尺寸（2.5, 2.1）较小，但花瓣特征对分类起决定性作用（SVM模型优先依赖花瓣数据）。

### 2. 数据验证：

- 数据集中类似样本（如第131行：**6.3,2.7,4.9,1.8,Iris-virginica**）的花瓣长度与输入值（4.9）一致，标签为 **virginica**。

### 3. 代码执行（可直接验证）：

python

复制

```
from sklearn.svm import SVC
import numpy as np

# 加载数据并训练模型（全数据集）
data = np.loadtxt('iris.txt', delimiter=',', dtype=str)
X = data[:, :4].astype(float)
y = data[:, 4]
model = SVC(kernel='linear').fit(X, y)

# 预测
features = np.array([[2.5, 2.1, 4.9, 1.5]])
print("预测结果:", model.predict(features)[0]) # 输出: Iris-virginica
```

请使用机器学习算法中的贝叶斯模型帮我预测一下，并直接告诉我答案。



**预测结果：** 这朵花的种类是 **Iris-virginica**（维吉尼亚鸢尾）。

## 关键说明

### 1. 模型选择：

- 使用高斯朴素贝叶斯（Gaussian Naive Bayes）模型，基于特征的条件概率分布进行分类。
- 贝叶斯模型对花瓣尺寸（4.9, 1.5）的敏感度较高，直接指向 **virginica** 类别。

### 2. 概率分析：

- 花瓣长度（4.9）远超 **versicolor** 的均值（约4.2），接近 **virginica** 的均值（约5.6），分类时概率显著偏向后者。

### 3. 代码验证：

```
python
复制
from sklearn.naive_bayes import GaussianNB
import numpy as np

# 加载数据
data = np.loadtxt('iris.txt', delimiter=',', dtype=str)
X = data[:, :4].astype(float)
y = data[:, 4]

# 训练贝叶斯模型
model = GaussianNB().fit(X, y)

# 预测
features = np.array([[2.5, 2.1, 4.9, 1.5]])
print("预测结果:", model.predict(features)[0]) # 输出: Iris-virginica
```

**注意：** 若花萼尺寸异常小（如2.5），但贝叶斯模型仍优先依赖花瓣特征，结论不变。

# 04

## 总结展望



# 总结

## ■ PPT制作

Deepseek + Kimi (三种工作模式)

## ■ 会议纪要

Deepseek + 通义听悟

会议视频/录音  $\xrightarrow{\text{AI转写工具}}$  会议文档  $\xrightarrow{\text{Deepseek}}$  结构化的会议纪要

## ■ 数据处理

数据生成、分析、标记、清洗、可视化、推理



## 展望

### ■ 短期：

- 熟练使用AI工具：毕业季的论文答辩、求职
- 建立“AI工具库”：辅助未来工作和科研。

### ■ 长期：

- 培养AI思维：遇到问题时先思考“AI能否解决”；保持对新技术的好奇心。
- 伦理与风险意识：警惕数据隐私、生成内容的真实性（如Deepfake风险）。



# 谢谢大家!

主讲人：朱一亨

南京农业大学 人工智能学院

2025年04月12日