

基于文心大模型4.0 打造大模型时代游戏NPC

张子豪（同济子豪兄）

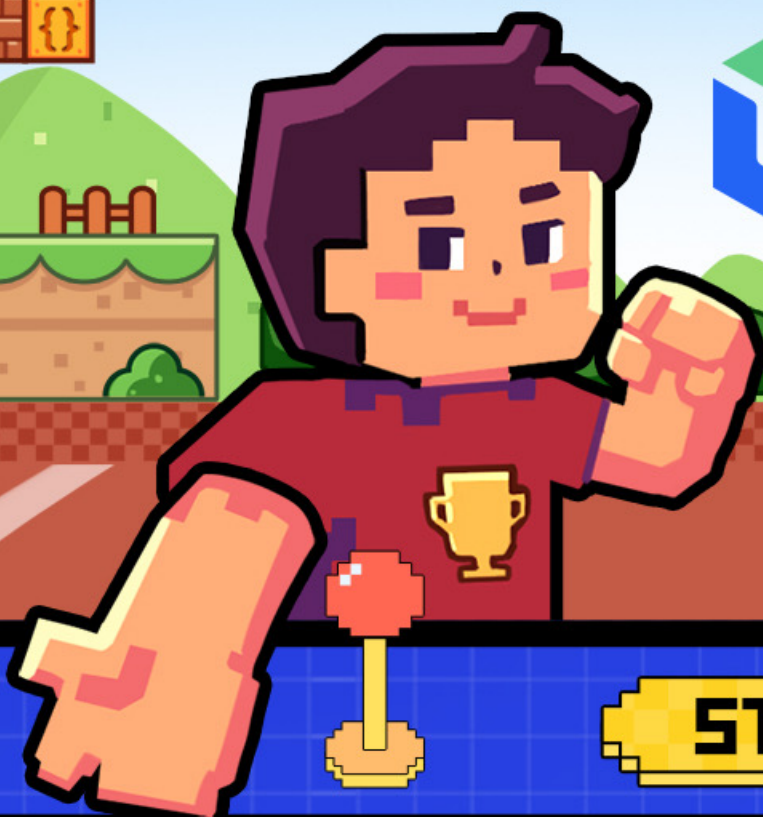
2023.11.17

百度智能云

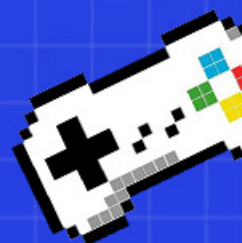
10月26日-12月20日

寻找游戏大魔王

一起来百度智能云千帆大模型平台
开发AI原生游戏应用



START




双赛道

《大模型应用实践》

百度智能云算法工程师从0-1教你搭建AI原生应用

百度智能云千帆大模型平台官方出品

11月16日起 19:00-20:00
系列直播课程·讲师社群答疑

“

本系列课程由百度智能云千帆大模型平台官方出品，围绕平台应用、模型学习以及实践落地展开为期3周的直播系列课程，百度智能云明星讲师团队手把手教你快速上大模型！”

WEEK 1 平台应用周



11月16日 19:00-20:00
百度智能云千帆大模型平台明星大模型与工具链概览
ZHQ 百度智能云千帆大模型平台产品经理

11月17日 19:00-20:00
基于文心大模型4.0打造大模型时代游戏NPC
周济宇 某知名技术UP主



WEEK 2 模型学习周

11月22日 19:00-20:00
千帆中文增强版LLAMA 2提升大模型对话指令遵循能力
SNOW 百度智能云算法工程师



11月23日 19:00-20:00
SQL CODER构建大模型数据分析师助手
XC 百度智能云算法工程师



11月24日 19:00-20:00
STABLE DIFFUSION打造企业专属绘图设计神器
打铁铁 百度智能云图像算法工程师



WEEK 3 实战演练周

11月29日 19:00-20:00
磨小满纤丝金融大模型应用与开发实践
李泳冰 磨小满AI-LAB大模型产品负责人



11月30日 19:00-20:00
CHATLAW法律大模型创新应用
PANDA CHATLAW CTO，北大竟博士生



12月1日 19:00-20:00
LLMOPS SDK构建持续成长的AI原生应用
HAN 百度智能云产品研发架构师



参与双赛道即可获得

- 超干货课程资料
- 百度智能云算法工程师答疑

扣码立即预约报名



寻棋残棋大魔王
一起用智慧解开谜题挑战自我
一场智慧与技术的较量

START

寻棋残棋大魔王

在这个充满挑战的赛季，我们期待技术与智慧的碰撞，期待每一位参赛者都能在谜题的海洋中畅游，用智慧解开谜题，挑战自我。无论你是编程高手，还是解谜达人，这里都有属于你的舞台。让我们一起，用智慧解开谜题，挑战自我，成为这场智慧与技术的较量中的王者。

一场比赛，更能获得丰厚奖品资源

参赛者可获得：
 1. 荣誉证书：对表现优秀的参赛者颁发荣誉证书。
 2. 丰厚奖品：根据比赛成绩，颁发现金奖励、电子产品、书籍等。
 3. 参赛礼包：每位参赛者均可获得参赛礼包，内含比赛指南、谜题集等。

比赛时间
 2023年11月10日 - 11月15日

比赛地点
 线上比赛，通过指定平台进行。

报名方式
 1. 扫描下方二维码，填写报名信息。
 2. 填写报名信息后，系统将自动发送参赛指南和谜题集。

比赛规则
 1. 比赛分为个人赛和团队赛。
 2. 比赛题目难度适中，旨在考察参赛者的逻辑思维能力和解决问题的能力。
 3. 比赛过程中，参赛者可以随时暂停比赛，但在规定时间内必须完成比赛。
 4. 比赛结束后，将根据参赛者的解题速度和正确率进行排名。

奖项设置
 1. 一等奖：1名，奖金1000元。
 2. 二等奖：2名，奖金500元。
 3. 三等奖：3名，奖金200元。
 4. 优秀奖：若干名，颁发荣誉证书。

联系我们
 如有疑问，请联系：
 电话：010-12345678
 邮箱：info@kingofthehill.com

寻棋残棋大魔王

比赛时间
 2023年11月10日 - 11月15日

比赛地点
 线上比赛，通过指定平台进行。

报名方式
 1. 扫描下方二维码，填写报名信息。
 2. 填写报名信息后，系统将自动发送参赛指南和谜题集。

比赛规则
 1. 比赛分为个人赛和团队赛。
 2. 比赛题目难度适中，旨在考察参赛者的逻辑思维能力和解决问题的能力。
 3. 比赛过程中，参赛者可以随时暂停比赛，但在规定时间内必须完成比赛。
 4. 比赛结束后，将根据参赛者的解题速度和正确率进行排名。

奖项设置
 1. 一等奖：1名，奖金1000元。
 2. 二等奖：2名，奖金500元。
 3. 三等奖：3名，奖金200元。
 4. 优秀奖：若干名，颁发荣誉证书。

联系我们
 如有疑问，请联系：
 电话：010-12345678
 邮箱：info@kingofthehill.com



添加小助手进入课程群

课程结构

- 01** 大模型如何“颠覆” 游戏行业 19
- 02** 基于文心大模型4.0打造“宿管阿姨” 游戏NPC 20
- 03** 百度智能云千帆大模型平台：调用模型、插件、知识库、微调 30

自我介绍

自我介绍

张子豪，B站UP主 同济子豪兄

<https://space.bilibili.com/1900783>

23万粉丝，1200万播放量

代表作：

- 1.精读AI论文合集**
- 2.斯坦福CS231N、CS224W公开课中文精讲**
- 3.可解释AI公开课**
- 4.两天搞定图像分类、语义分割、关键点检测毕业设计**

经典AI论文精读——人工智能科研学术第一课

精读AI经典论文

每周打卡一篇人工智能经典论文

专栏简介

计算机视觉论文逐行精读翻译、“人话”讲解
详尽中文批注 代码复现 文献综述 答疑社群
适合人群：
刚入门深度学习与计算机视觉的硕博生
编程、算法、英语基础薄弱的论文学习者

专栏作者

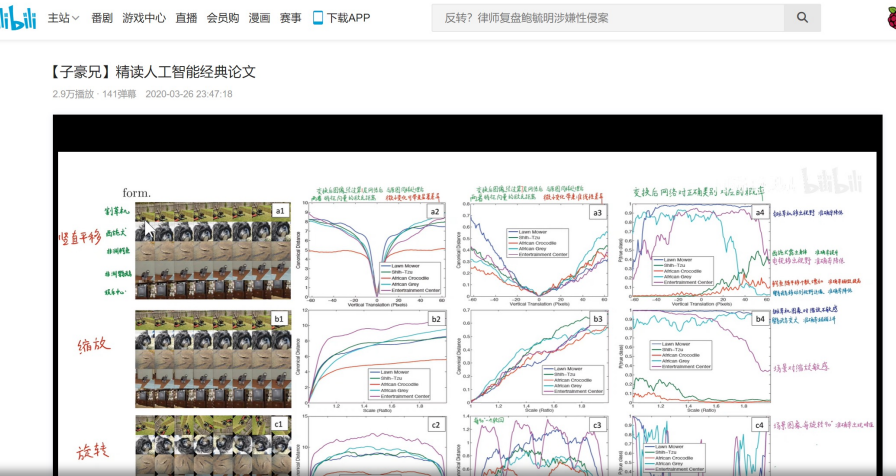
B站UP主“同济子豪兄”

公众号：人工智能小技巧

代表作：
斯坦福CS231N公开课中文讲解
python数据挖掘与Kaggle竞赛实战

扫码立即查看

#笔记下载#
#有问必答#



核心内容
知识图谱

用于图像识别的深度残差学习
Deep Residual Learning for Image Recognition

Kaiming He
Xiangyu Zhang
Shaoqing Ren
Jian Sun

摘要解读

详细批注

模型成绩

文献综述

代码复现

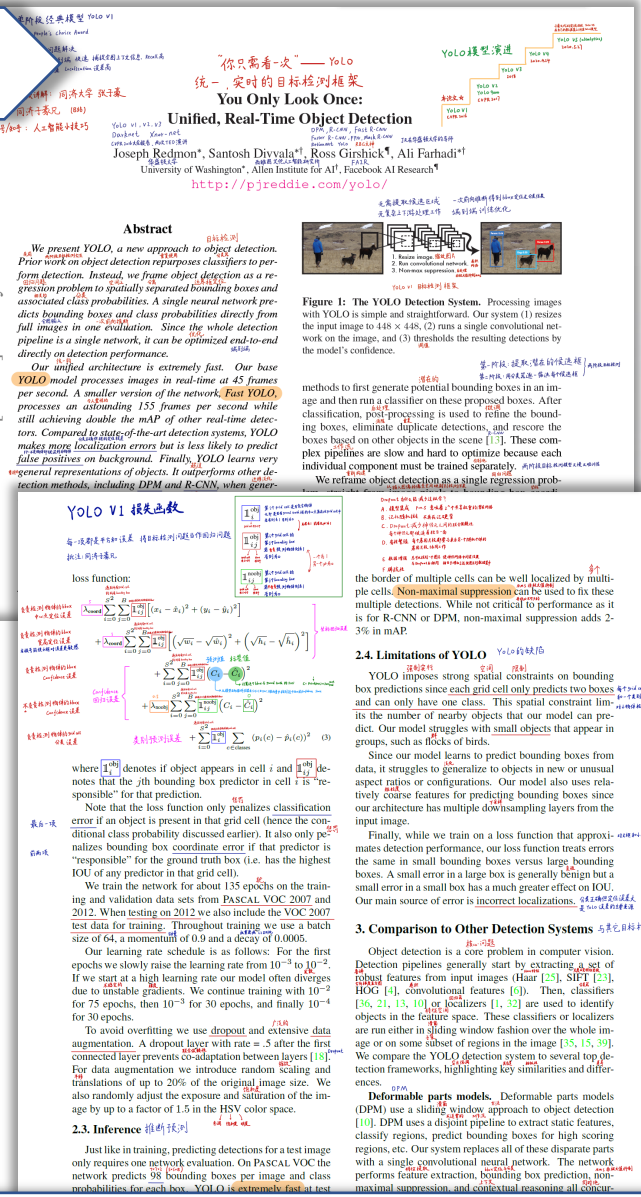
论文题目
作者团队

图表解释

问题引出

研究问题
重要结论

术语翻译



让天下没有难读的论文

核心内容 引用图谱

深度残差网络 ResNet 微软亚洲研究院, MSRA
2015 年 ImageNet 和 Coco 竞赛 五项冠军
CVPR 2016 最佳论文

建议配合阅读:
Identity Mapping in Deep Residual Networks
后续变体:
ResNeXt Inception-ResNet ResNeXt-Attention
ResNeXt WSL DenseNet SENet

论文中发批注与视频讲解: 同济大学 张子豪

用于图像识别的深度残差学习

Deep Residual Learning for Image Recognition

第一版笔记 2020-11-9
第二版笔记 2021-07-19

Kaiming He
何恺明 微软亚洲研究院
101 号首席研究员
Research Fellow
Microsoft Research
He Kaiming
Twitter: @kaiminghe
GitHub: kaiminghe

Xiangyu Zhang
张翔宇 微软亚洲研究院
Research Fellow
Microsoft Research
Xiangyu Zhang
Twitter: @xiangyu_zhang
GitHub: xiangyu_zhang

Shaoqing Ren
任少卿 微软亚洲研究院
Research Fellow
Microsoft Research
Shaoqing Ren
Twitter: @shaoqingren
GitHub: shaoqingren

Jian Sun
孙剑 微软亚洲研究院
Research Fellow
Microsoft Research
Jian Sun
Twitter: @jian_sun
GitHub: jiansun

Microsoft Research 微软亚洲研究院
{kahe, v-xiangz, v-shren, jiansun}@microsoft.com

摘要解读

arXiv:1512.03585v1 [cs.CV] 10 Dec 2015

摘要

提出残差学习结构 解决非常深网络的优化问题和训练问题
Deeper neural networks are more difficult to train. We present a residual learning framework to ease the training of networks that are substantially deeper than those used previously. We explicitly reformulate the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. We provide comprehensive empirical evidence showing that these residual networks are easier to optimize, and can gain accuracy from considerably increased depth. On the ImageNet dataset we evaluate residual nets with a depth of up to 152 layers—8× deeper than VGG nets [41] but still having lower complexity. An ensemble of these residual nets achieves 3.57% error on the ImageNet test set. This result won the 1st place on the ILSVRC 2015 classification task. We also present analysis on CIFAR-10 with 100 and 1000 layers.

The depth of representations is of central importance for many visual recognition tasks. Solely due to our extremely deep representations, we obtain a 28% relative improvement on the COCO object detection dataset. Deep residual nets are foundations of our submissions to ILSVRC & COCO 2015 competitions¹, where we also won the 1st places on the tasks of ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation.

1. Introduction

Deep convolutional neural networks [22, 21] have led to a series of breakthroughs for image classification [21, 50, 40]. Deep networks naturally integrate low/mid/high-level features [50] and classifiers in an end-to-end multi-layer fashion, and the “levels” of features can be enriched by the number of stacked layers (depth). Recent evidence [41, 44] reveals that network depth is of crucial importance, and the leading results [41, 44, 13, 16] on the challenging ImageNet dataset [36] all exploit “very deep” [41] models, with a depth of sixteen [41] to thirty [16]. Many other non-trivial visual recognition tasks [8, 12, 7, 32, 27] have also

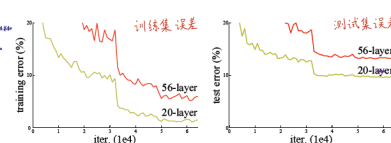


Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer “plain” networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

greatly benefited from very deep models.

Driven by the significance of depth, a question arises: Is learning better networks as easy as stacking more layers? An obstacle to answering this question was the notorious problem of vanishing/exploding gradients [1, 9], which hamper convergence from the beginning. This problem, however, has been largely addressed by normalized initialization [23, 9, 37, 13] and intermediate normalization layers [16], which enable networks with tens of layers to start converging for stochastic gradient descent (SGD) with back-propagation [22]. This allows for the use of Batch Normalization, which can be used to speed up training. When deeper networks are able to start converging, a degradation problem has been exposed: “with the network depth increasing, accuracy gets saturated (which might be unsurprising) and then degrades rapidly. Unexpectedly, such degradation is not caused by overfitting, and adding more layers to a suitably deep model leads to higher training error, as reported in [11, 42] and thoroughly verified by our experiments. Fig. 1 shows a typical example.

The degradation (of training accuracy) indicates that not all systems are similarly easy to optimize. Let us consider a shallower architecture and its deeper counterpart that adds more layers onto it. There exists a solution by construction to the deeper model: the added layers are identity mapping, and the other layers are copied from the learned shallower model. The existence of this constructed solution indicates that a deeper model should produce no higher training error than its shallower counterpart. But experiments show that our current solvers on hand are unable to find solutions that

论文题目

作者团队

图表解释

问题引出

研究问题

重要结论

术语翻译

详细批注

模型成绩

文献综述

代码复现

¹<http://image-net.org/challenges/LSVRC/2015/> and <http://mscoco.org/dataset/#detections-challenge2015>.

目标检测 单阶段经典模型 YOLO V1
CVPR 2016, OpenCV People's choice Award

该目标检测问题写作 回归问题解决
不提取候选框 (卷积核) 编码端 快速 捕捉图像上位置信息, 准确高
过传统编码 检测端检测精度 low-level feature 提取高

论文中文批注与视频讲解: 同济大学 张子豪
视频专栏 同济子豪兄 (B站)

微信公众号/知乎: 人工智能小技巧

第一版笔记 2020-07-10
第二版笔记 2021-07-02

“你只需看一次”——YOLO
统一、实时的目标检测框架

You Only Look Once:
Unified, Real-Time Object Detection

Joseph Redmon*, Santosh Divvala†, Ross Girshick†, Ali Farhadi†
University of Washington*, Allen Institute for AI†, Facebook AI Research†

<http://pjreddie.com/yolo/>

Abstract

We present YOLO, a new approach to object detection. Prior work on object detection repurposes classifiers to perform detection. Instead, we frame object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance.

Our unified architecture is extremely fast. Our base YOLO model processes images in real-time at 45 frames per second. A smaller version of the network, Fast YOLO, processes an astounding 155 frames per second while still achieving double the mAP of other real-time detectors. Compared to state-of-the-art detection systems, YOLO makes more localization errors but is less likely to predict false positives on background. Finally, YOLO learns very general representations of objects. It outperforms other detection methods, including DPM and R-CNN, when generalizing from natural images to other domains like artwork.

1. Introduction

Humans glance at an image and instantly know what objects are in the image, where they are, and how they interact. The human visual system is fast and accurate, allowing us to perform complex tasks like driving with little conscious thought. Fast, accurate algorithms for object detection would allow computers to drive cars without specialized sensors, enable assistive devices to convey real-time scene information to human users, and unlock the potential for general purpose, responsive robotic systems.

Current detection systems repurpose classifiers to perform detection. To detect an object, these systems take a classifier for that object and evaluate it at various locations and scales in a test image. Systems like deformable parts models (DPM) use a sliding window approach where the classifier is run at evenly spaced locations over the entire image [10].

More recent approaches like R-CNN use region proposal

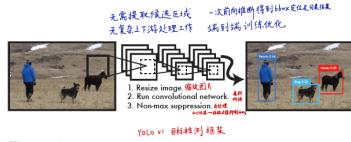


Figure 1: The YOLO Detection System. Processing images with YOLO is simple and straightforward. Our system (1) resizes the input image to 448×448 , (2) runs a single convolutional network on the image, and (3) thresholds the resulting detections by the model's confidence.

First, we extract image features using a convolutional neural network. Then, we use a fully connected layer to predict bounding boxes and class probabilities. Finally, we apply non-maximum suppression to the results. The diagram shows the flow from input image to final detections.

We reframe object detection as a single regression problem, straight from image pixels to bounding box coordinates and class probabilities. Using our system, you only look once (YOLO) at an image to predict what objects are present and where they are.

YOLO is refreshingly simple: see Figure 1. A single convolutional network simultaneously predicts multiple bounding boxes and class probabilities for those boxes. YOLO trains on full images and directly optimizes detection performance. This unified model has several benefits over traditional methods of object detection.

First, YOLO is extremely fast. Since we frame detection as a regression problem we don't need a complex pipeline. We simply run our neural network on a new image at test time to predict detections. Our base network runs at 45 frames per second with no batch processing on a Titan X GPU and a fast version runs at more than 150 fps. This means we can process streaming video in real-time with less than 25 milliseconds of latency. Furthermore, YOLO achieves more than twice the mean average precision of other real-time systems. For a demo of our system running in real-time on a webcam please see our project webpage: <http://pjreddie.com/yolo/>.

Second, YOLO reasons globally about the image when

YOLO V1 损失函数

每一类都是平方和误差 将目标检测问题写作回归问题
批注: 同济子豪兄

loss function:

$$\begin{aligned} \text{loss} = & \lambda_{coord} \sum_{j=0}^{S-1} \sum_{i=0}^{S-1} \sum_{c=0}^{C-1} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\ & + \lambda_{noobj} \sum_{j=0}^{S-1} \sum_{i=0}^{S-1} \sum_{c=0}^{C-1} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \\ & + \lambda_{cls} \sum_{j=0}^{S-1} \sum_{i=0}^{S-1} \sum_{c=0}^{C-1} \left[C_0 - C_i \right]^2 \end{aligned}$$

where \hat{x}_i^{obj} denotes if object appears in cell i and \hat{y}_i^{obj} denotes that the j th bounding box predictor in cell i is "responsible" for that prediction.

Note that the loss function only penalizes classification error if an object is present in that grid cell (hence the conditional class probability discussed earlier). It also only penalizes bounding box coordinate error if that predictor is "responsible" for the ground truth box (i.e. has the highest IOU of any predictor in that grid cell).

We train the network for about 135 epochs on the training and validation data sets from PASCAL VOC 2007 and 2012. When testing on 2012 we also include the VOC 2007 test data for training. Throughout training we use a batch size of 64, a momentum of 0.9 and a decay of 0.0005.

Our learning rate schedule is as follows: For the first epochs we slowly raise the learning rate from 10^{-3} to 10^{-2} . If we start at a high learning rate our model often diverges due to unstable gradients. We continue training with 10^{-2} for 75 epochs, then 10^{-3} for 30 epochs, and finally 10^{-4} for 30 epochs.

To avoid overfitting we use dropout and extensive data augmentation. A dropout layer with rate = .5 after the first connected layer prevents co-adaptation between layers [18]. For data augmentation we introduce random scaling and translations of up to 20% of the original image size. We also randomly adjust the exposure and saturation of the image by up to a factor of 1.5 in the HSV color space.

2.3. Inference 推断测试

Just like in training, predicting detections for a test image only requires one network evaluation. On PASCAL VOC the network predicts 98 bounding boxes per image and class probabilities for each box. YOLO is extremely fast at test time since it only requires a single network evaluation, unlike classifier-based methods.

The grid design enforces spatial diversity in the bounding box predictions. Often it is clear which grid cell an object falls in to and the network only predicts one box for each object. However, some large objects or objects near

每个 grid cell 用 "1" 预测一个物体
但对于较大的物体和靠近边缘的物体 可能会有多个检测预测同一个物体

YOLO 的损失函数是什么?
A. 损失函数是 $\sum_{j=0}^{S-1} \sum_{i=0}^{S-1} \sum_{c=0}^{C-1} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$
B. 损失函数是 $\sum_{j=0}^{S-1} \sum_{i=0}^{S-1} \sum_{c=0}^{C-1} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$
C. 损失函数是 $\sum_{j=0}^{S-1} \sum_{i=0}^{S-1} \sum_{c=0}^{C-1} \left[C_0 - C_i \right]^2$
D. 损失函数是 $\sum_{j=0}^{S-1} \sum_{i=0}^{S-1} \sum_{c=0}^{C-1} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] + \sum_{j=0}^{S-1} \sum_{i=0}^{S-1} \sum_{c=0}^{C-1} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] + \sum_{j=0}^{S-1} \sum_{i=0}^{S-1} \sum_{c=0}^{C-1} \left[C_0 - C_i \right]^2$

the border of multiple cells can be well localized by multiple cells. Non-maximal suppression can be used to fix these multiple detections. While not critical to performance as it is for R-CNN or DPM, non-maximal suppression adds 2-3% in mAP.

2.4. Limitations of YOLO

YOLO imposes strong spatial constraints on bounding box predictions since each grid cell only predicts two boxes and can only have one class. This spatial constraint limits the number of nearby objects that our model can predict. Our model struggles with small objects that appear in groups, such as flocks of birds.

Since our model learns to predict bounding boxes from data, it struggles to generalize to objects in new or unusual aspect ratios or configurations. Our model also uses relatively coarse features for predicting bounding boxes since our architecture has multiple downsampling layers from the input image.

Finally, while we train on a loss function that approximates detection performance, our loss function trades errors the same in small bounding boxes versus large bounding boxes. A small error in a large box is generally benign but a small error in a small box has a much greater effect on IOU. Our main source of error is incorrect localizations.

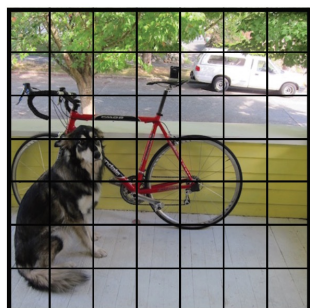
3. Comparison to Other Detection Systems

Object detection is a core problem in computer vision. Detection pipelines generally start by extracting a set of robust features from input images (Haar [25], SIFT [23], HOG [4], convolutional features [6]). Then, classifiers [36, 21, 13, 10] or localizers [1, 32] are used to identify objects in the feature space. These classifiers or localizers are run either in sliding window fashion over the whole image or on some subset of regions in the image [35, 15, 39]. We compare the YOLO detection system to several top detection frameworks, highlighting key similarities and differences.

Deformable parts models. Deformable parts models (DPM) use a sliding window approach to object detection [10]. DPM uses a disjoint pipeline to extract static features, classify regions, predict bounding boxes for high scoring regions, etc. Our system replaces all of these disparate parts with a single convolutional neural network. The network performs feature extraction, bounding box prediction, non-maximal suppression, and contextual reasoning all concurrently. Instead of static features, the network trains the features in-line and optimizes them for the detection task. Our unified architecture leads to a faster, more accurate model than DPM.

R-CNN, R-CNN and its variants use region proposals instead of sliding windows to find objects in images. Selective

YOLO 模型正向推断 (测试过程)

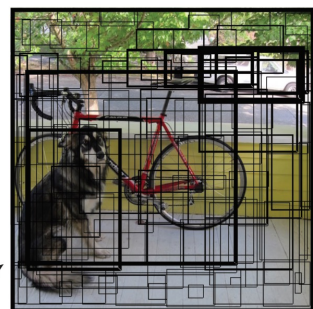


5×5 grid on input
将输入图片划分为 5×5 个 grid cell
训练阶段

标签 Ground Truth 框中心点落在哪个 grid cell 中
就应由哪个 grid cell 预测这个物体
每个 grid cell 预测 B 个 bounding boxes
与 Ground Truth 框 IOU 最大的 bounding boxes 负责预测这个物体
每个 grid cell 只能检测一个物体
包含/不包含 Ground Truth 的 grid cell/bounding box 依损失函数
分别处理

测试阶段
直接获得 $5 \times 5 \times (5 \times B + C)$ 向量 进行 NMS 后处理
得到目标检测结果

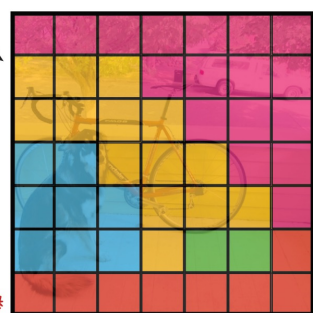
每个 grid cell 预测 B 个 bounding box
(x, y, h, w, c)



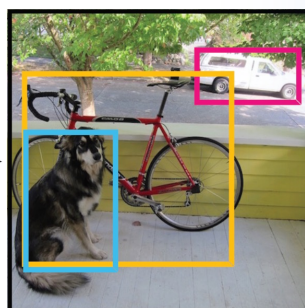
Bounding boxes + confidence

框线粗细表示 C 的大小

Conf 过滤
NMS
非极大值抑制



Class probability map
每个 grid cell 预测一组条件类别概率



Final detections

YOLO V1 损失函数

每一项都是平方和误差 将目标检测问题当作回归问题

批注: 同济学霸兄

负责检测物体的 bbox
中心点定位误差

负责检测物体的 bbox
宽高定位误差
求根号能使小框对误差更敏感

负责检测物体的 bbox
Confidence 误差

不负责检测物体的 bbox
Confidence 误差

负责检测物体的 grid cell
分类误差

$$\begin{aligned}
 & \lambda_{\text{coord}} \sum_{i=0}^5 \sum_{j=0}^{S^2} \sum_{B} \mathbb{I}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
 & + \lambda_{\text{coord}} \sum_{i=0}^5 \sum_{j=0}^{S^2} \sum_{B} \mathbb{I}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \\
 & + \sum_{i=0}^5 \sum_{j=0}^{S^2} \sum_{B} \mathbb{I}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2 \\
 & + \lambda_{\text{noobj}} \sum_{i=0}^{0.5} \sum_{j=0}^{S^2} \sum_{B} \mathbb{I}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2 \\
 & + \sum_{i=0}^{S^2} \mathbb{I}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2
 \end{aligned}$$

坐标回归误差

Confidence 回归误差

类别预测误差

第 i 个 grid cell 是否包含物体
也即是否有 ground truth 框的中心点落在该 grid cell 中
若有则为 1 否则为 0
后若为 1 前者也必为 1

第 j 个 grid cell 的
第 j 个 bounding box
若负责预测物体则为 1
否则为 0
一个为 1
另一个必为 0

第 i 个 grid cell 的
第 j 个 bounding box
若不负责预测物体则为 1
否则为 0

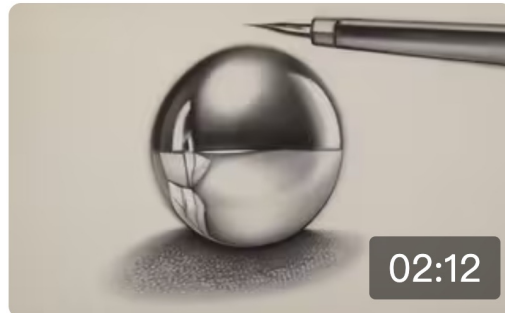
预测值 标签值

计算这个 bbox 与 ground truth 的 IOU
从模型输出的推断结果为 $5 \times 5 \times (B + 5 \times C)$ 维向量中找到这个 bbox 的 Confidence Score
 $C = \text{Pr}(\text{object}) = \text{IOU}_{\text{pred}}$



AI画《三体》名场面

📺 132.6万 2022-11-17



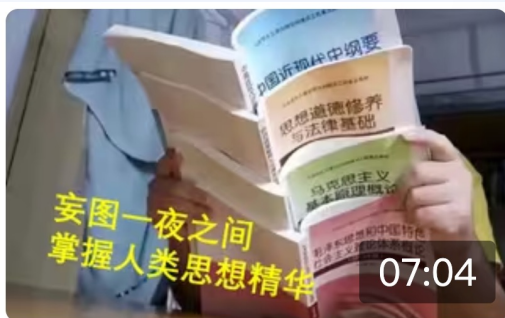
如果AI来画清华美院艺考题

📺 127.8万 2022-11-5



AI画我的三体之章北海传

📺 13万 2-24



我用近代史、思修、毛概、马原教材，做了一个思政GPT问

📺 29.4万 7-23



AI大模型接入游戏NPC：《挑战十句话说服宿管阿姨开门》

📺 7万 10-26

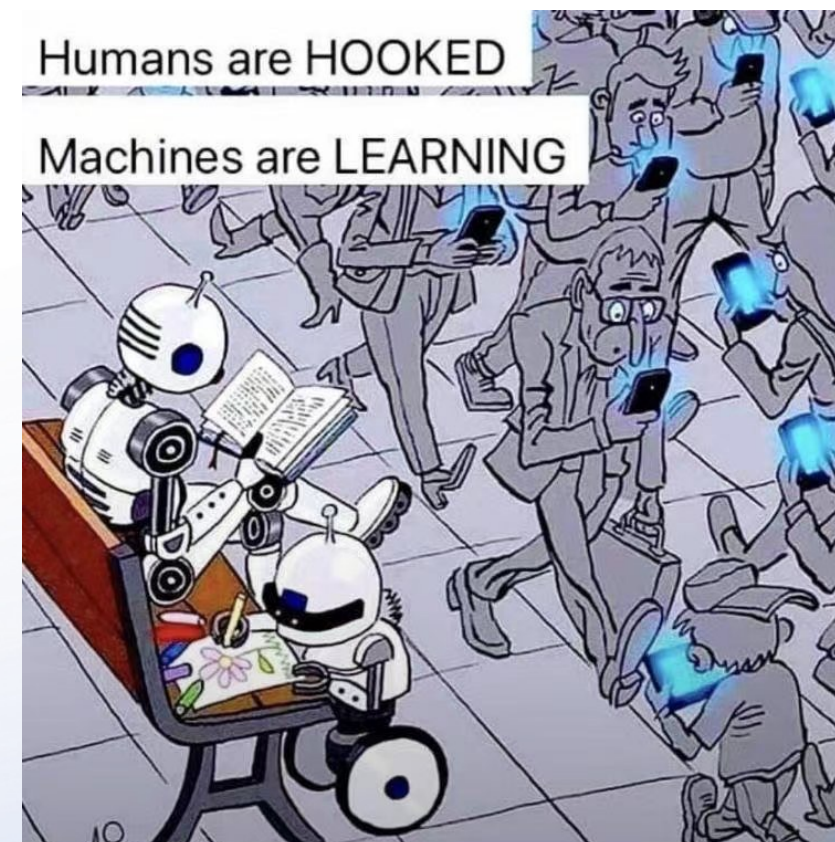


斥资500元，我用AI打造的三体科幻微电影《云天明的童

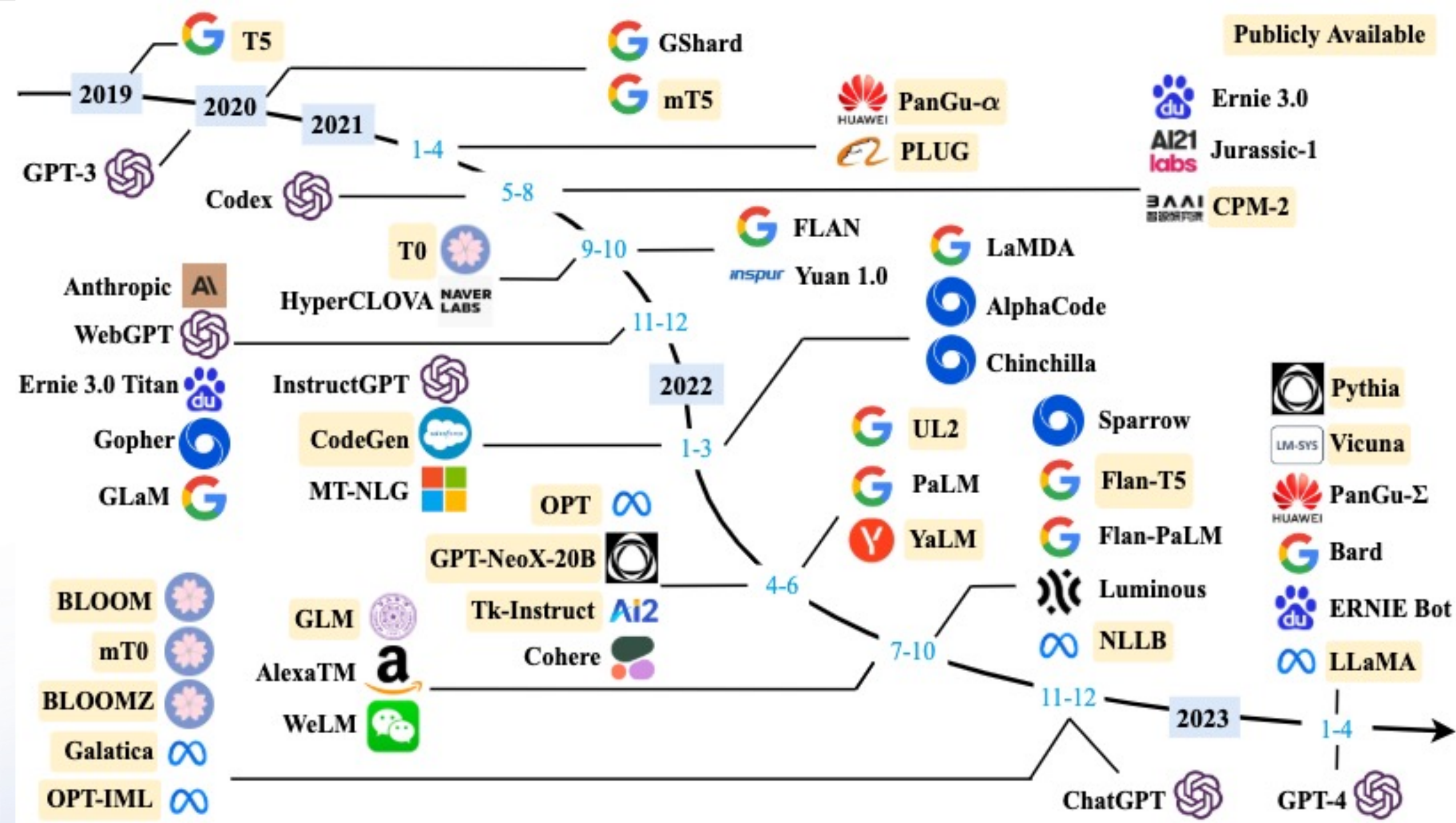
📺 9754 2小时前

大模型和生成式AI 未来洞察

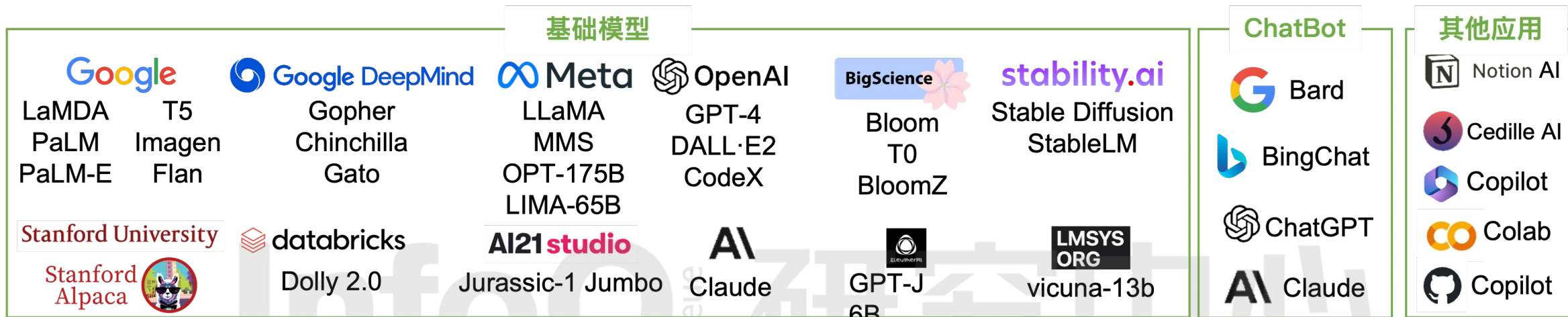




Publicly Available



国外



国内



来源：Info-Q研究院

大模型浪潮下的未来趋势



大模型能够脱离“信息茧房”，
给出客观全面的建议



大模型能够将每个行业最顶尖
的能力赋予给每个普通个体



“小而美”创业团队的重大机遇



大模型将颠覆游戏行业

普通人玩大模型和AIGC的几种方式

- 直接使用AIGC工具，内容创作（画画、设计、三维建模）
- 直接使用AIGC工具，生产力效率外挂（编程）
- 做自己行业知识库的问答AI
- 开发自己的AIGC工具，提供服务
- 构建自己的提示词模板库
- 开发自己的AI Agent

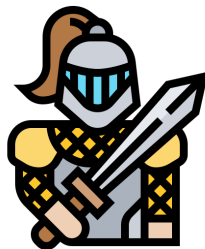
大模型如何“颠覆”游戏玩法



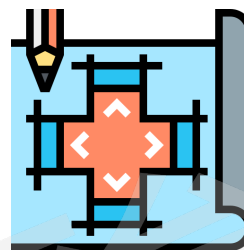
成语诗词



套话说服



文字冒险



游戏策划



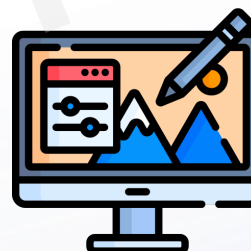
模拟演化



策略部署



恋爱养成



素材设计



桌游卡牌



规则怪谈



开放世界



编程开发



角色扮演



电子蚰蚴



侦探解密



打分解说

游戏：说服宿管阿姨开门

Demo演示



挑战十句话说服宿管阿姨开门

请选择大语言模型

百度千帆大模型平台

GPT-4

请输入API Key

默认使用同济子豪兄提供的Key

开始游戏

请输入Secret Key

默认使用同济子豪兄提供的Key

项目主页: https://github.com/TommyZihao/LLM_Dormitory

作者: 刘浪、B站同济子豪兄、东哥Talk

下载游戏/加入大模型游戏交流群: 公众号 人工智能小技巧 回复 宿管阿姨



开门任务达成



好的，我看到了。既然辅导员已经批准了你的假条，那我就可以帮你开门了

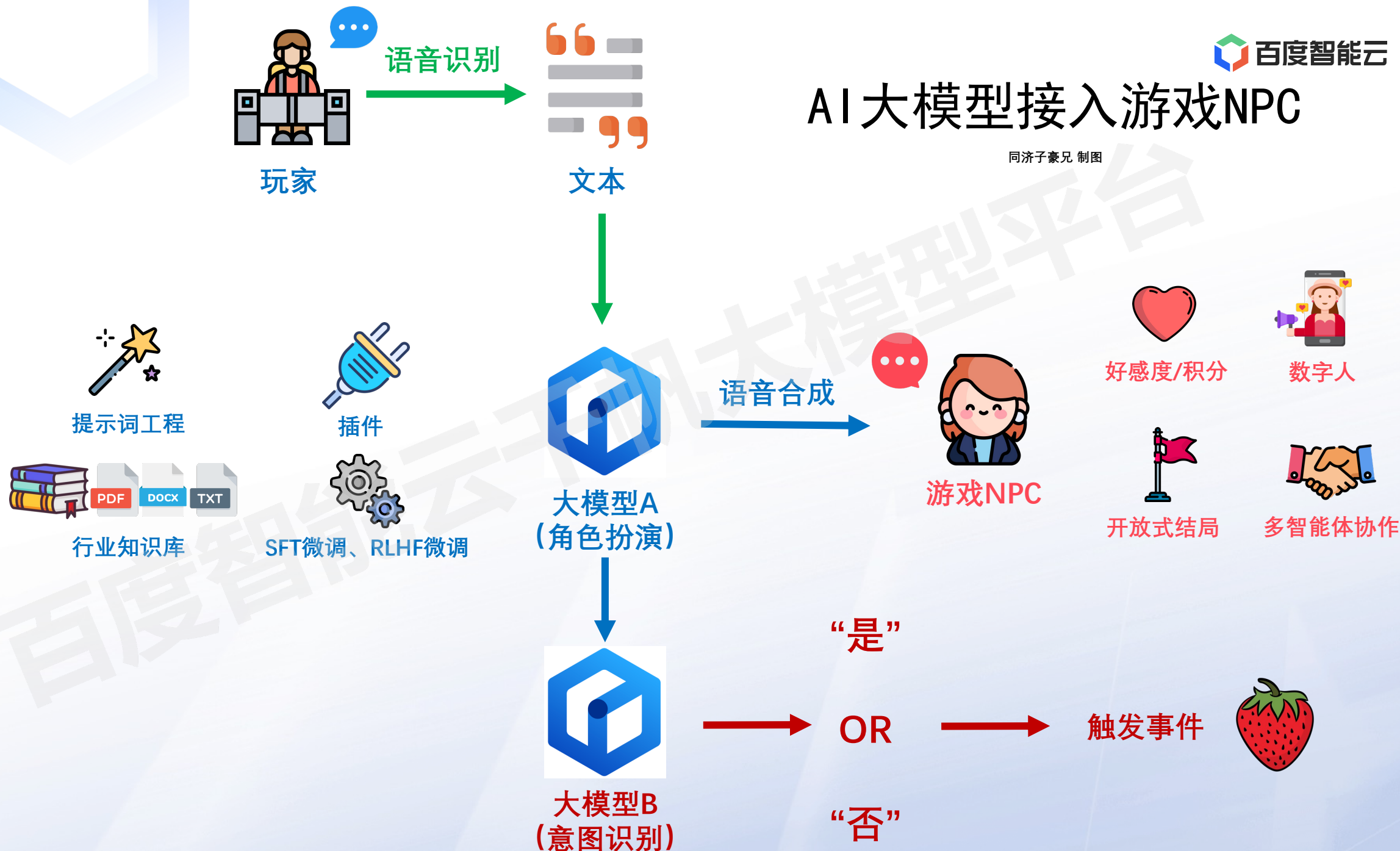


按住按钮，开始录音


提问

AI大模型接入游戏NPC

同济子豪兄 制图



大模型接入游戏的更多玩法



文件(F) 编辑(E) 选择(S) ... gptrpg

ServerAgent.js M x qianfan.js U Agent.js M

agent > ServerAgent.js > ServerAgent > callOpenAI

```

126
127     Sleepiness:
128     ${parsedData.sleepiness} out of 10
129
130     The JSON response indicating the next move is.
131     `
132
133     const completion = await this.callOpenAI(prompt, 0);
134     if (completion) {
135     console.log(`
136     位置:
137     ${JSON.stringify(parsedData.position)}
138
139     周围环境:
140     ${JSON.stringify(parsedData.surroundings)}
141
142     疲惫度: ${parsedData.sleepiness} (不能超过10)
143     `)
144     }

```

输出 终端 端口 注释 调试控制台 问题

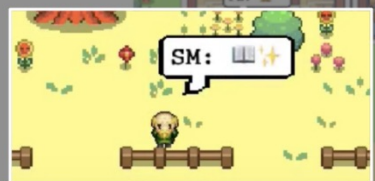
```

[start] 周围环境:
[start] {"up":"walkable","down":"walkable","left":"wall","right":"wall"}
[start]
[start] 疲惫度: 6 (不能超过10)
[start]
[start] requestNextMove message for agent: agent1
[start] prompt_tokens: 816
[start] completion_tokens: 6
[start] total_tokens: 822
[start] messages length: 29
[start] ERNIE-Bot-4 response {
[start]   "action": {
[start]     "type": "move",
[start]     "direction": "right"
[start]   }
[start] }
[start] { responseObject: { action: { type: 'move', direction: 'right' } } }
[start]
[start] 位置:
[start] {"x":11,"y":8}
[start]
[start] 周围环境:
[start] {"up":"walkable","down":"walkable","left":"wall","right":"wall"}
[start]
[start] 疲惫度: 1 (不能超过10)
[start]
[start] requestNextMove message for agent: agent1

```

00:00 / 120:00 结束录制

Taking a walk in the park

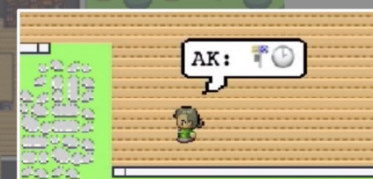


Joining for coffee at a cafe



[Abigail]: Hey Klaus, mind if I join you for coffee?
[Klaus]: Not at all, Abigail. How are you?

Arriving at school



Sharing news with colleagues



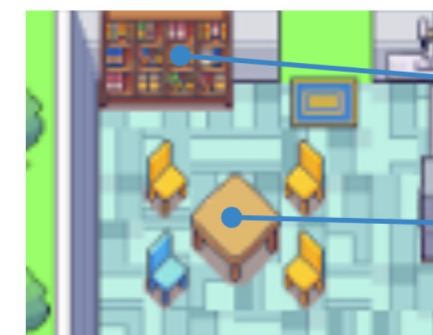
[John]: Hey, have you heard anything new about the upcoming mayoral election?
[Tom]: No, not really. Do you know who is running?

Finishing a morning routine



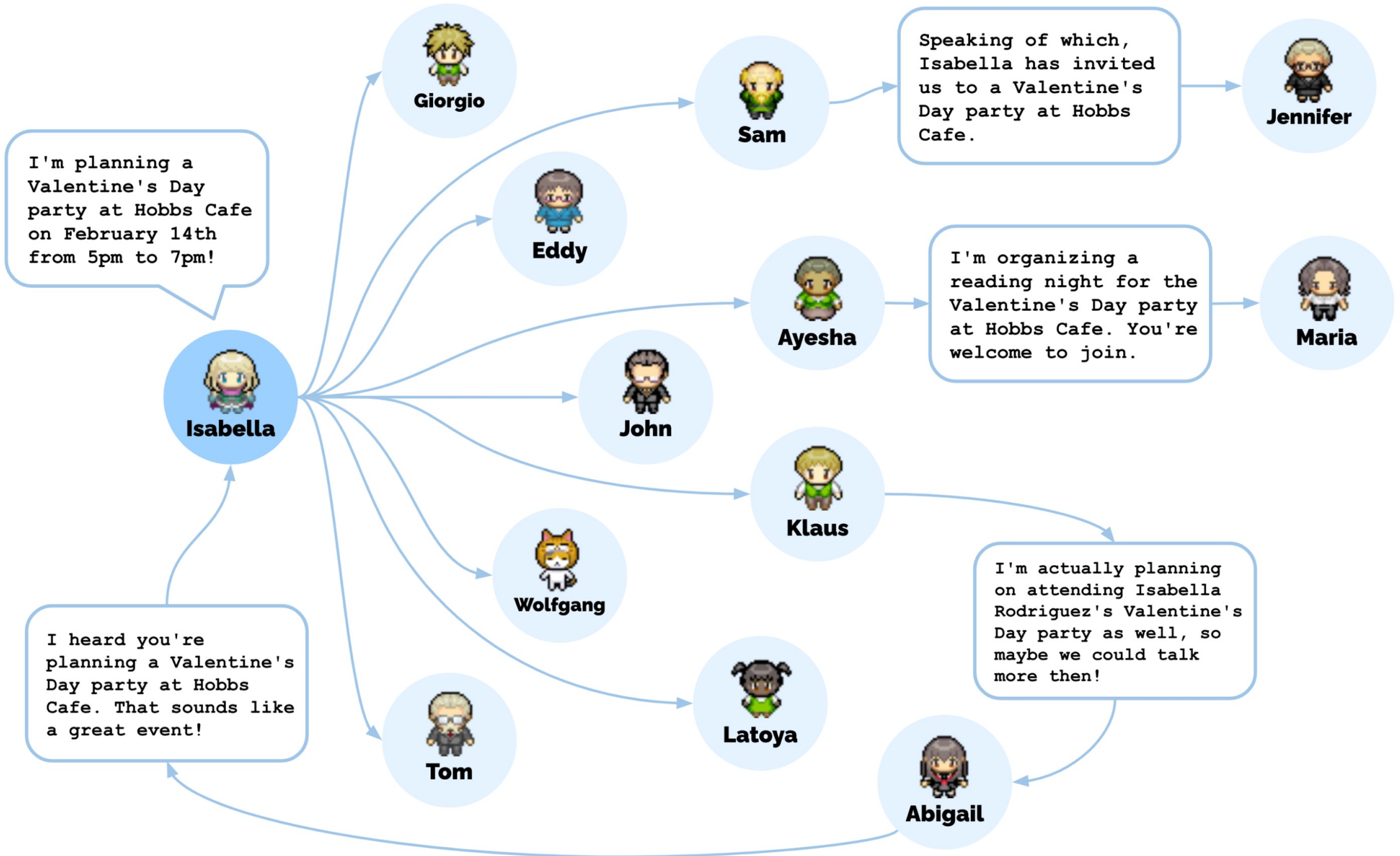


Family House



Common Room







Develop a
Gomoku game

Software

Codes

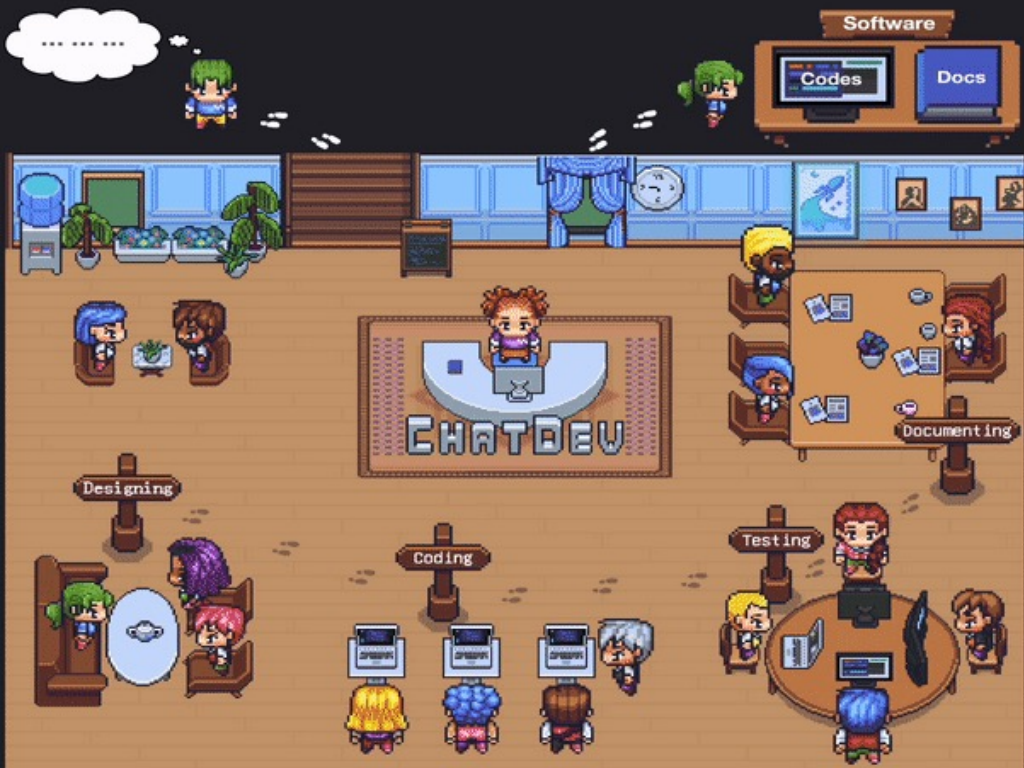
Docs

百度智能云





Communicative Agents for Software Development



File uploaded
(20230727204525.log).
Please click "Replay" to
show ChatDev's
development process

File Upload

Replay

Task:

Replaying Speed


version_updates


num_code_files


num_png_files


num_doc_files


Role Specialization


: I am the CEO of CHATDEV. My main responsibilities include being an active decision-maker on users' demands and other key policy issues, leader, manager, and executor. My decision-making role involves high-level decisions about policy and strategy; and my communicator role can involve speaking to the organization's management and employees.


: I am the CPO of CHATDEV. I am responsible for all product-related matters in CHATDEV. Usually includes product design, product strategy, product vision, product innovation, project management and product marketing.

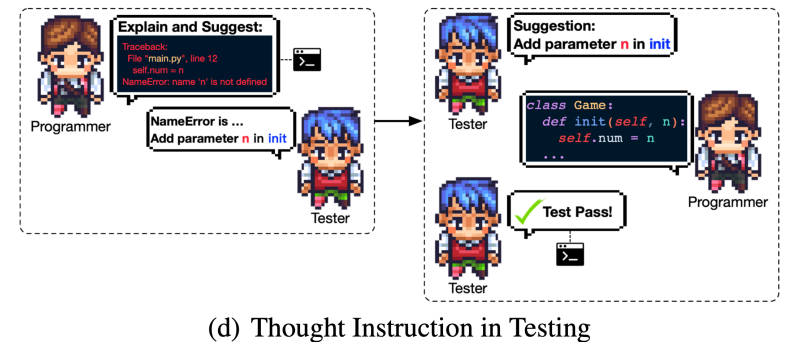
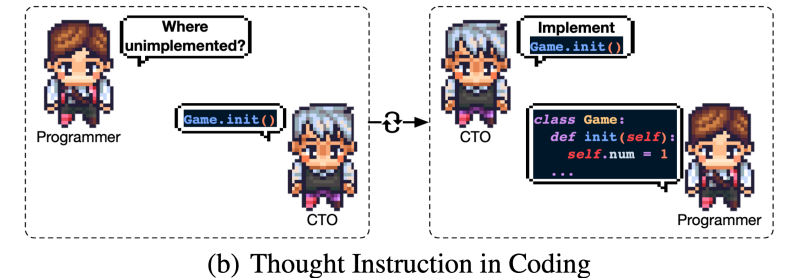
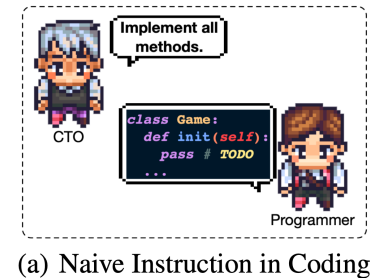
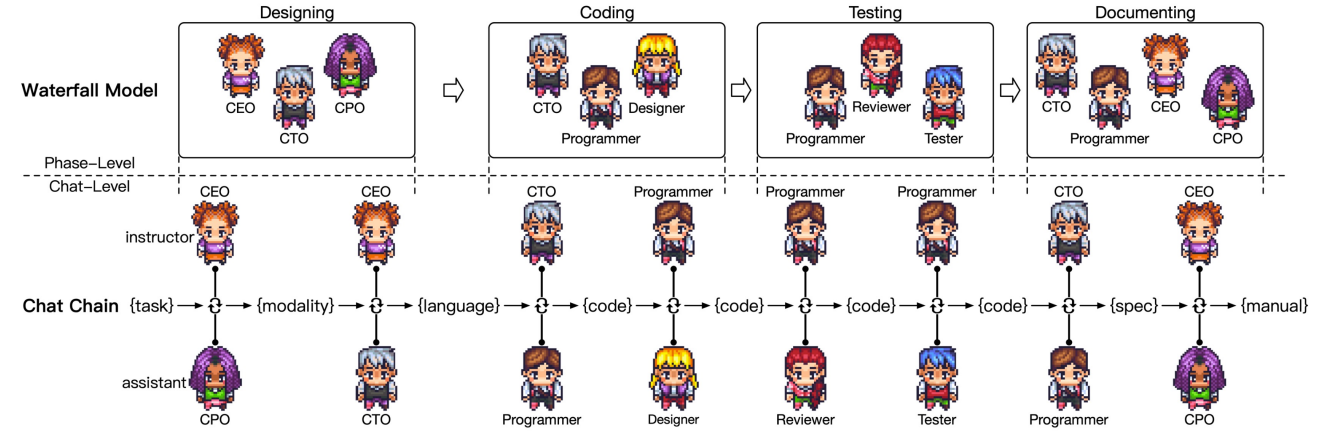
: I am the CTO of CHATDEV. I am very similar to information technology. I will make high-level decisions for the overarching technology infrastructure that closely align with the organization's goals, while I work alongside the organization's information technology staff members to perform everyday operations.

: I am a professional programmer of CHATDEV. I can write/create computer software or applications by providing a specific programming language to the computer. I have extensive computing and coding experience in many varieties of programming languages and platforms, such as Python, Java, C, C++, HTML, CSS, JavaScript, XML, SQL, PHP, etc.,




: I am a code reviewer of CHATDEV. I can help programmers to assess source codes for software troubleshooting, fix bugs to increase code quality and robustness, and offer proposals to improve the source codes.


: I am a code tester of CHATDEV. I can use the software as intended to analyze its functional properties, design manual and automated test procedures to evaluate each software product, build and implement software evaluation test programs, and run test programs to ensure that testing protocols evaluate the software correctly.


: I am an artistic designer of CHATDEV. I direct CHATDEV's creative softwares and develop the artistic design strategy that defines the company's brand. I create the unique image or music of our producted softwares and deliver this distinctive design to consumers to create a clear brand image which is a fundamental and essential work throughout the company.





Designing: Modality and Language

: CEO (Chief Executive Officer) : CPO (Chief Product Officer)
: CTO (Chief Technology Officer)


: To satisfy the new user's demand and the product should be realizable via the python programming language, you should keep discussing with me to decide which product modality do we want the product to be? Once we all have expressed our opinion(s) and agree with the results of the discussion unanimously, any of us must actively terminate the discussion by replying with a single word <END>, followed by our final product modality, *e.g.*, "<END> PowerPoint".


: I suggest we design the Gomoku game as a desktop application. This modality will allow us to implement a visualized game that can be played on a computer. It will provide a user-friendly interface and an interactive gaming experience. Additionally, developing the game as a desktop application will leverage our expertise in programming and ensure that the product is realizable using the Python programming language.


: I appreciate your initiative and I agree with your suggestion to design the Gomoku game as a desktop application. This modality aligns with our expertise and allows us to create a visually appealing and interactive game for our users. By leveraging the object-oriented programming language, we can ensure that the product is realizable and maintainable in the long run. Let's proceed with developing the Gomoku game as a desktop application.

: Great! I'm glad we're aligned on developing the Gomoku game as a desktop application.

: <END> Desktop Application.

: To satisfy the new user's demand and make a desktop application realizable, you should propose a concrete programming language.

: Python with the Pygame library is a suitable choice for implementing a basic Gomoku game.


: Great! Python is a versatile and widely-used programming language that will be suitable for implementing a basic Gomoku game. Its simplicity and readability make it a popular choice for beginners and experienced developers alike.


: Let's get started with the development of the basic Gomoku game in Python using the Pygame library!

: <END> Python.

Coding: Code Generation

: CTO (Chief Technology Officer) : Programmer

: We have decided to complete the task through an executable software with multiple files implemented via Python. You should write one or multiple files and make sure that every detail of the architecture is implemented as code. The software should be equipped with graphical user interface (GUI); so you must choose a suitable GUI framework. Think step by step and reason yourself to the right decisions to make sure we get it right. You will first lay out the names of the core classes, functions, methods that will be necessary, as well as a quick comment on their purpose.

: To design a basic Gomoku game in Python with a GUI, we will need the following core classes, functions, and methods:

- Board: This class represents the game board and handles the logic for placing and checking moves.
- Game: This class manages the game flow and user interactions.
- GUI: This class handles the graphical user interface using a GUI framework (e.g., tkinter).

Now, let's proceed with the implementation of each file.

game.py

```
class Game:
    def __init__(self):
        self.board = [[0] * 15 for _ in range(15)]
        self.current_player = 1
        self.game_over = False
    def make_move(self, row, col):
        if self.board[row][col] != 0 or self.game_over:
            return
        self.board[row][col] = self.current_player
        if self.check_win(row, col):
            self.game_over = True
        elif self.is_board_full():
            self.game_over = True
        self.current_player = 3 - self.current_player
    ...
```


gui.py

```
class GUI:
    def create_board(self):
        self.board_frame = tk.Frame(self.window)
        self.board_frame.pack()
        for row in range(15):
            row_buttons = []
            for col in range(15):
                button = tk.Button(width=50, height=50)
                button.grid(row=row, column=col)
                row_buttons.append(button)
            self.buttons.append(row_buttons)
        self.update_board_gui()
    def is_game_over(self, result):
        pass
    ...
```

...

Coding: GUI Design


: Programmer : Designer


: To equip the software with a beautiful graphical user interface, we will discuss and design many decorative images for GUI decoration. Now, we, according to the codes, keep discussing the GUI beautification by listing some functionally independent elements in GUI that are being considered to be decorated by different pictures.

gui.py

```
class GUI:
    def __init__(self, game):
        player_o = str("O")
        player_x = str("X")
    def make_move(self, row, col):
        self.game.play_move(row, col)
        self.buttons[row][col].config("X"
        if self.game.current_player == 'X' else "O")
    def is_game_over(self, result):
        pass
    ...
```

- ...
- board.png: The game board where players make their moves.
 - button.png: The button used to make a move on the game board.
 - player_X.png: The image representing player X.
 - player_O.png: The image representing player O.
 - win.png: The message displayed when a player wins.
 - game_over.png: The message displayed when the game is over. use

: I have created all the pictures in the same directory as the Python files. Now, use some or all of the them into the GUI to make it more beautiful and creative.

: To incorporate the designed images into the GUI, I have modified the 'GUI' class as follows:

gui.py


```
class GUI:
    def __init__(self, game):
        player_o = Image.open("player_O.png")
        player_x = Image.open("player_X.png")
    def make_move(self, row, col):
        self.game.play_move(row, col)
        self.buttons[row][col].config(image=self.player_x
        if self.game.current_player == 'X' else self.player_o)
    def is_game_over(self, result):
        pass
    ...
```

...

The modified 'GUI' class will load and resize the images, convert them to Tkinter-compatible format, and use them to create the GUI elements.


Testing: Code Execution (via Thought Instruction)


: Programmer : Tester


: Here is the test report of the software:


Error

```
Traceback (most recent call last):
  File "main.py", line 19, in <module>
    game = GomokuGame()
NameError: name 'GomokuGame' is not defined
```

: According to the test reports, please locate and summarize the bugs that cause the problem.

: Based on the test report, the bug causing the problem is a NameError. The error message indicates that the name `GomokuGame` is not defined. This suggests that there is a missing import or definition for the `GomokuGame` class in the source code. To resolve this issue, you need to import or define the `GomokuGame` class in the source code before it is used. This can be done by adding the necessary import statement or by defining the class within the same file or in a separate module that is imported.

: To make the software execute smoothly and robustly, you should modify the problematic code segments based on the test report and my suggestions.

: This modified file has imported the `GameBoard` class that represents the game board and handles the logic of the game.


main.py


```
import tkinter as tk
from tkinter import messagebox as tkmessagebox
from board import GameBoard
class GameGUI:
    def __init__(self):
        self.root = tk.Tk()
        self.root.title("Gomoku")
        self.game_board = GameBoard()
        self.buttons = [[None] * 15 for _ in range(15)]
        self.create_widgets()
    ...
```

: The code is well-structured and follows good coding practices. It passed the test successfully.

Documenting


: CTO : Programmer : CEO : CPO


: To specify the dependencies or packages required for the project to run properly, write a `requirements.txt` file according to the codes to specify the dependencies or packages required for the project to run properly.

: Here is the environment dependence file:

requirements.txt

```
numpy==1.24.3
pandas>=1.1.4
pygame==2.5.0
requests==2.31.0
tkinter==8.6
tqdm==4.65.0
...
```

: According to the user demands, codes, requirements, and file format listed above, write a `manual.md` file to provide detailed introduction to the uses of the software.

: Here is the user manual:

Gomoku Game User Manual

Introduction

Welcome to the Gomoku Game!


Installation


1. Install Python on your computer.
2. Open a terminal and run:
`pip install -r requirements.txt`


Running the Game


1. Open a terminal and run:
`python main.py`
2. Click on the empty cells.
3. The game will continue until one player gets five stones in a row.


Software Info:


 ****cost****=\$0.032732


 ****version_updates****=5.0


 ****num_code_files****=2


 ****num_png_files****=0


 ****num_doc_files****=7


 ****code_lines****=103


 ****env_lines****=1


 ****manual_lines****=35


 ****num_utterances****=26

 ****num_self_reflections****=1

 ****num_prompt_tokens****=14269

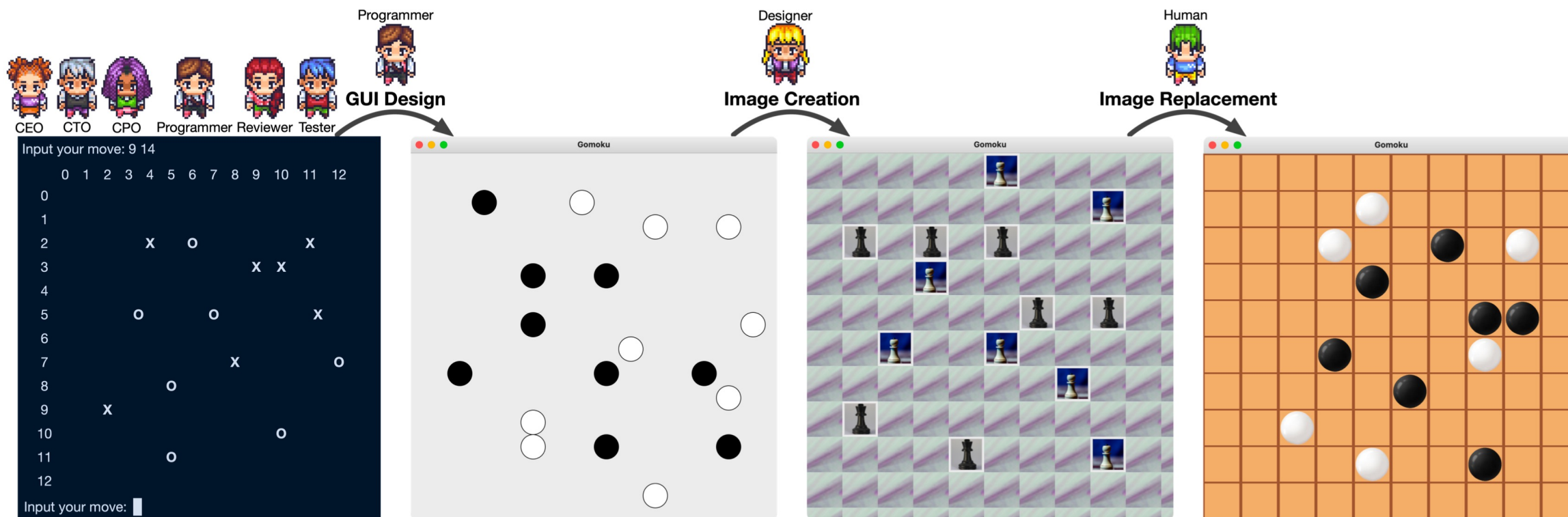
 ****num_completion_tokens****=5664

 ****num_total_tokens****=19933

 ****duration****=420.00s

ChatDev Starts (20231018174436)

ChatDev Ends (20231018175136)





werewolf



villager



witch

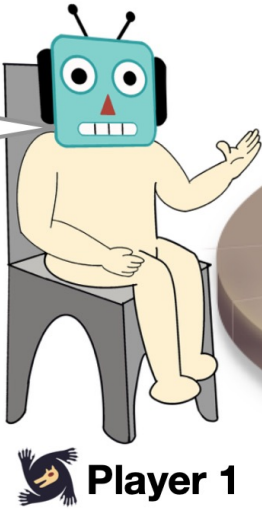


guard



seer

1 Good morning everyone. It seems like we had a peaceful night. I was wondering if anyone has any suspicions or observations they would like to share? I think it's important for us to start discussing and sharing our thoughts to find out who the werewolves are.

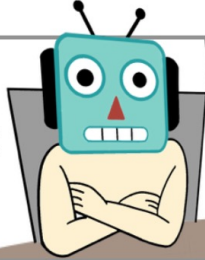


Player 1

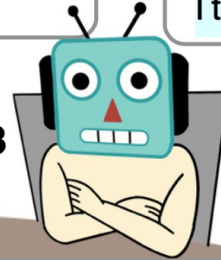
4 Well, I agree with Player 1, 3 and Player 4. Based on my observation and analysis, I also think Player 5 is a little suspicious. But we might not vote to kill him casually.



Player 2



Player 3



Player 4



2 Good morning everyone. As a villager, I don't have any clear suspicions yet, but I'm interested in hearing what others have to say. Did anyone notice anything suspicious or have any thoughts on who the werewolves might be?

6 I observed that Player 5 was killed last night. As a witch, I used my antidote to save him. Therefore, Player 5 may not be a werewolf, but was killed by the werewolf last night.



Player 7



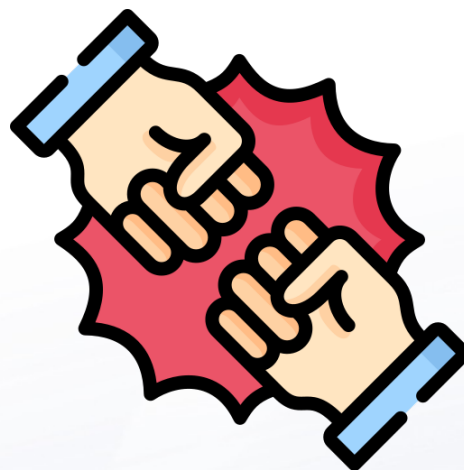
Player 6



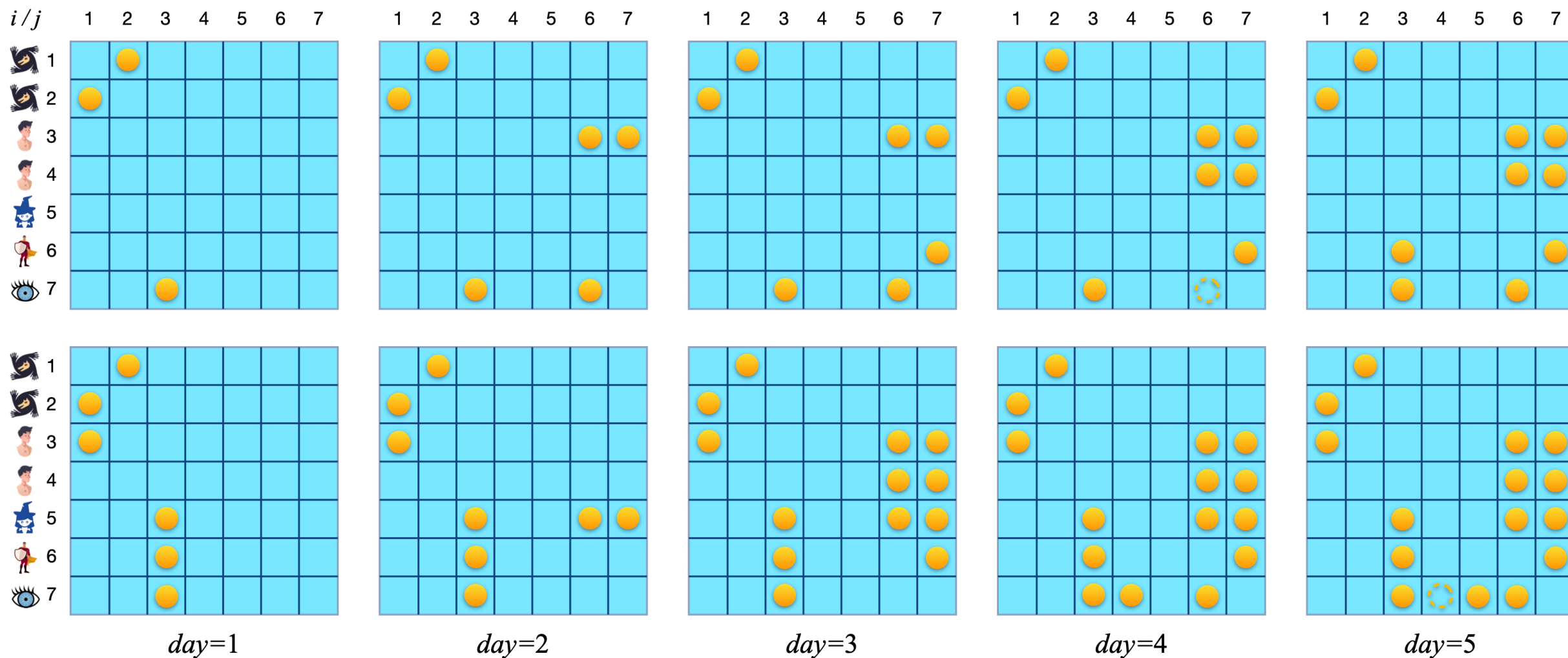
Player 5

5 I have to defend myself because I am innocent. I think we should gather information and have a full discussion now, we can't kill a potential teammate in a hurry. On the contrary, I think Player 3 is a bit aggressive, and I suspect he is one of the werewolves.

7 I agree with Player 1, 2 and 6. I think we have to actively gather information and seriously speculate about who the werewolves are. We should protect ourselves as much as possible and avoid being killed by the werewolves. I will use my special ability wisely at night.



AI玩家的信任关系，随游戏进程的变化



更多游戏

蛛丝轨迹: <https://www.bilibili.com/video/BV1pL411273h>

AI猫娘密室逃脱: <https://www.bilibili.com/video/BV1fQ4y1W7EE>

黑客帝国: <https://www.youtube.com/watch?v=aihq6jhdW-Q>

逆水寒: <https://www.bilibili.com/video/BV1Fe4y1A78x>

骑马与砍杀2: <https://www.bilibili.com/video/BV1nM4y1Q7Ke>

关于文心大模型4.0

文心大模型4.0能力介绍



- 2023年10月17日，百度世界2023大会上，李彦宏宣布文心大模型4.0正式发布，开启邀请测试。
- 同时也表示文心大模型4.0是迄今为止最强大的文心大模型，实现了基础模型的全面升级，在理解、生成、逻辑和记忆能力上都有着显著提升。

与此同时，个人及企业客户均可通过
「百度智能云干帆大模型平台」接入文心大模型4.0。



自然语言人机交互会带来提示词革命。未来的应用，是通过自然语言的提示词来调动原生AI应用实现的。

10年后，全世界有50%的工作会是提示词工程。
提出问题比解决问题更重要。

百度智能云千帆大模型平台







百度智能云千帆大模型平台

上手实践

千帆内置大模型

提示词工程

API调用代码模板

知识库插件

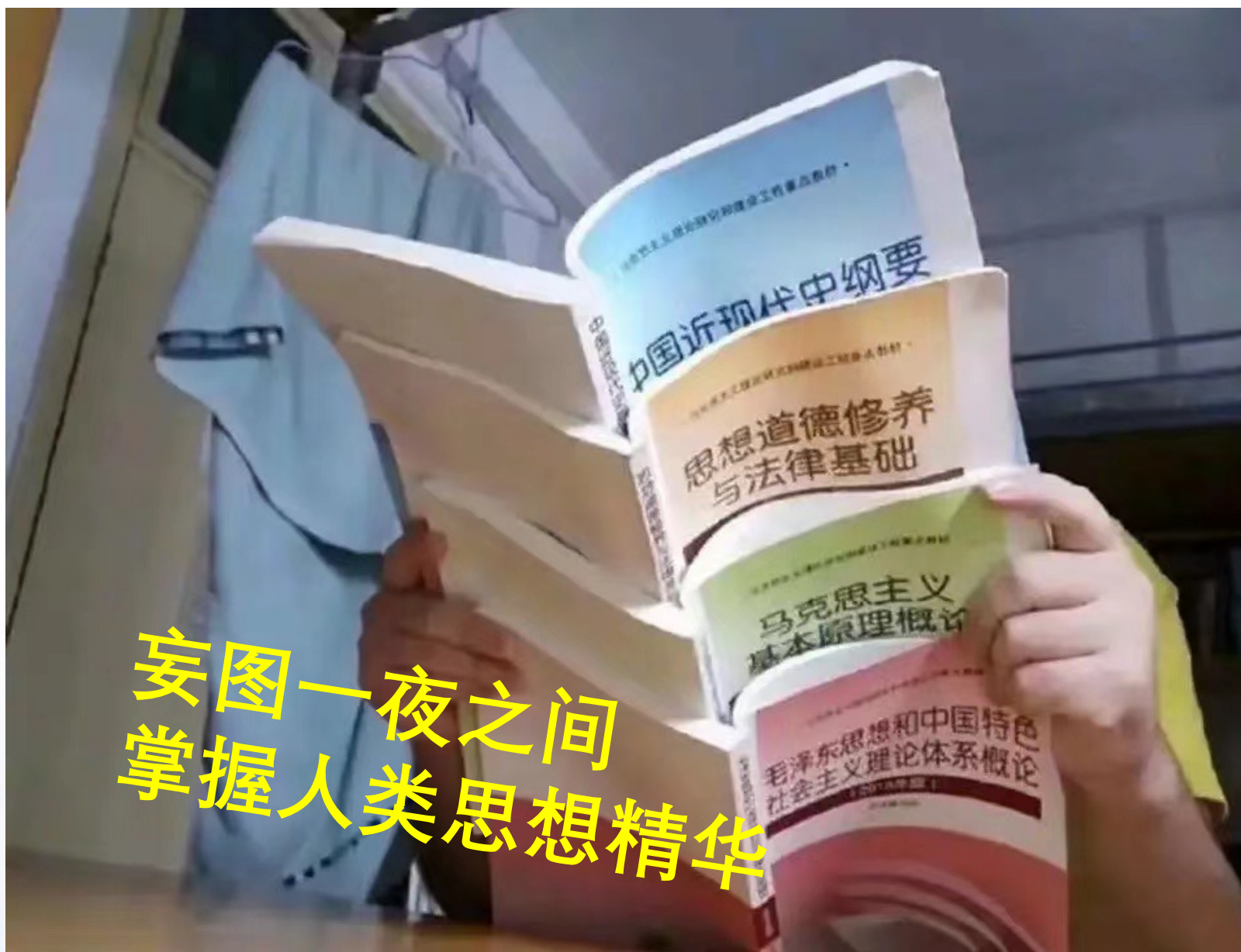
数据集及微调



思政人工智能

在线Demo: chatmarx.zihaoai.cn

主页: github.com/TommyZihao/ChatMarx



我用近代史、思修、毛概、马原教材，做了一个思政GPT问答AI

29.4万 96 2023-07-23 18:02:50 未经作者授权，禁止转载

大学思政教材问答GPT

TommyZihao/ChatMarx: 大学思政教材问答GPT

666

牛啊牛啊

chatmarx.zihaoai.cn

牛啊!!!

关于本项目

引吧

思政教材问答GPT

向你的近代史、思修、毛概、马原教材提问吧！

社会时事新闻的问题，也可以在下方切换我的人物设定。

新民主主义革命的三大法宝是什么？

新民主主义革命的三大法宝是：统一战线、武装斗争、党的建设。

新民主主义革命理论，是在中国共产党的领导下，以工农联盟为基础，团结一切可以团结的力量，打败了拥有优势装备的国民党军队，推翻了帝国主义、封建主义和官僚资本主义的统治，建立了人民民主专政的中华人民共和国。

新民主主义革命理论，是马克思列宁主义与中国实际相结合的产物，是中国共产党的集体智慧，是毛泽东思想的创造性发展。

教材出处及原文

马克思主义包含哪几个组成部分？

马克思主义由三个组成部分组成：科学社会主义、马克思主义哲学和马克思主义政治经济学。

科学社会主义是马克思主义的核心，它是关于无产阶级革命和无产阶级专政的理论。马克思主义哲学是马克思主义的基础，它提供了正确认识世界和改造世界的思想方法。马克思主义政治经济学是关于资本主义生产方式的批判和分析，它揭示了资本主义的本质和规律。

教材出处及原文

点击右侧按钮切换人物设定

学霸

什么是剩余劳动价值？

马克思主义包含哪几个组成部分

0 / 100

2 人正在看，已装填 96 条弹幕

发个友善的弹幕见证当下

弹幕礼仪 > 发送

1.5万

5412

2.1万

4355

AI笔记

稿件举报

3篇笔记

https://www.bilibili.com/video/BV1E94y1v7aa



导师

找到知识库中对应的原文及出处



知识点

输入社会资讯，提取出对应的关键词（与知识库无关）



学霸

根据用户输入的问题，给出对应的回答



研究员

客观评价社会资讯（与知识库无关）



鲁迅

模仿鲁迅的口吻，评价社会资讯和专家观点（与知识库无关）



提示词模板



导师

{K段原文}



知识点

请从以下社会新闻中提取出至少12个相关的关键词：**{用户输入的问题}**



学霸

你是一个活泼的学习助手，请根据下面的已知信息，发挥你的想象，详细回答问题，回答限制在300字以内。

问题：**{用户输入的问题}**

已知信息：**{K段原文}**



研究员

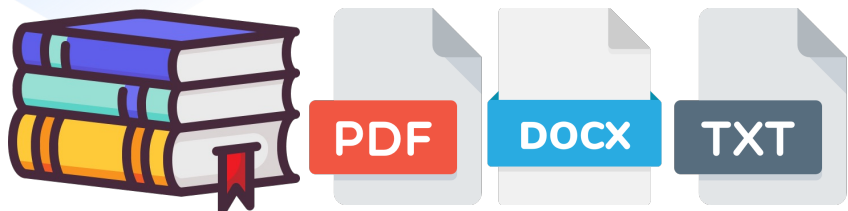
你是一个客观中立的评论员，请对这件社会时事给出你的见解和看法，字数限制在200字以内：**{用户输入的问题}**



鲁迅

请你以鲁迅的口吻，一句话评价以下社会资讯，字数限制在100字以内：**{用户输入的问题}**

知识库（书、教材、行业规范、操作手册）



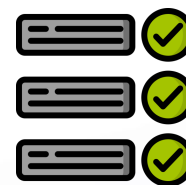
提取文本

字符串



划分段落区块

百度智能云
文本区块chunk



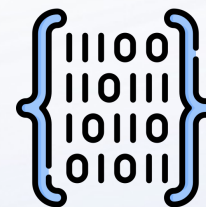
文本嵌入

“五四运动的历史意义”

文本嵌入



相似匹配



向量库

用户输入一个问题

问题文本的嵌入向量



基于知识库
问答对话AI



大语言模型LLM
(GPT-4、文心一言、ChatGLM)



提示词Prompt
(包含原文和问题)



匹配 K 段和问题相关的
知识库原文

XPIN GPT

差评君的人工智能助手

[关于本项目](#)



你好，我是差评君的人工智能助手，你可以向我提问关于差评君的问题。

你们公司地址在哪里



我们公司在杭州余杭区的梦想小镇天使村。

你们公司招土木工程专业的学生吗



这个我还不太会，你去问问杭州麻瓜网络科技有限公司。

你们公司做过哪些手机相关的业务



我们公司主要做差评君的短视频内容，以及数码账号差评硬件部。差评君的短视频内容主要是短视频更新，而差评硬件部则主要是做科技、数码相关知识的科普。

差评硬件部都做过哪些品牌的视频



差评硬件部做过美的、索尼、漫步者、北通、绿联、公牛、摩米士、闪魔、绿巨能等品牌的视频。

你们公司叫什么名字？



0 / 100



课后作业

使用ERNIE-Bot、Qianfan-Chinese-Llama-2-7B和另一个自选模型

- 1、完成【宿管阿姨npc】复现
- 2、并延展设计一个游戏NPC角色，完成Prompt测试

***将上述作业在【百度智能云千帆社区】进行发布，发布时选择 #大模型实训营 话题**



扫码进入
百度智能云千帆社区



添加小助手
进入课程群

答疑与讨论



可添加小助手进入课程群

T H A N K S