



A wide-angle aerial photograph of the Shanghai skyline during sunset. The city is densely packed with skyscrapers, many of which are illuminated with warm lights. In the foreground, the Huangpu River flows through the city. On the right side, the Oriental Pearl Tower stands prominently, its distinctive spherical observation deck glowing. The Bund area is visible in the background, with its iconic buildings and bridges reflected in the water. The sky is a mix of deep blues and warm orange and yellow hues from the setting sun.

DeepLaw

专属深度法律顾问

问题背景



案件 多

2015年全国法院新收
案件近 1800万 件，
同比上升 22.81%

法制网 2016.03.18



法律 细

我国现行法律法规中，
共有 29 类大大小小
共计 2569 部

中国企业集成网数据



咨询 难

我国执业律师人数已超
过 29.7万 人，律师权
务所达到 2.4万 多家

新华网 2016.03.30

法律咨询需求 与日剧增，却不知 找谁帮忙、从何问起，而且 费用高、耗时多

应用场景

DeepLaw

专注于 法律 垂直领域的人工智能聊天机器人



随便唠唠

一般聊天机器人所具备的
日常交流能力
可调侃、可逗趣、可玩耍



法律咨询

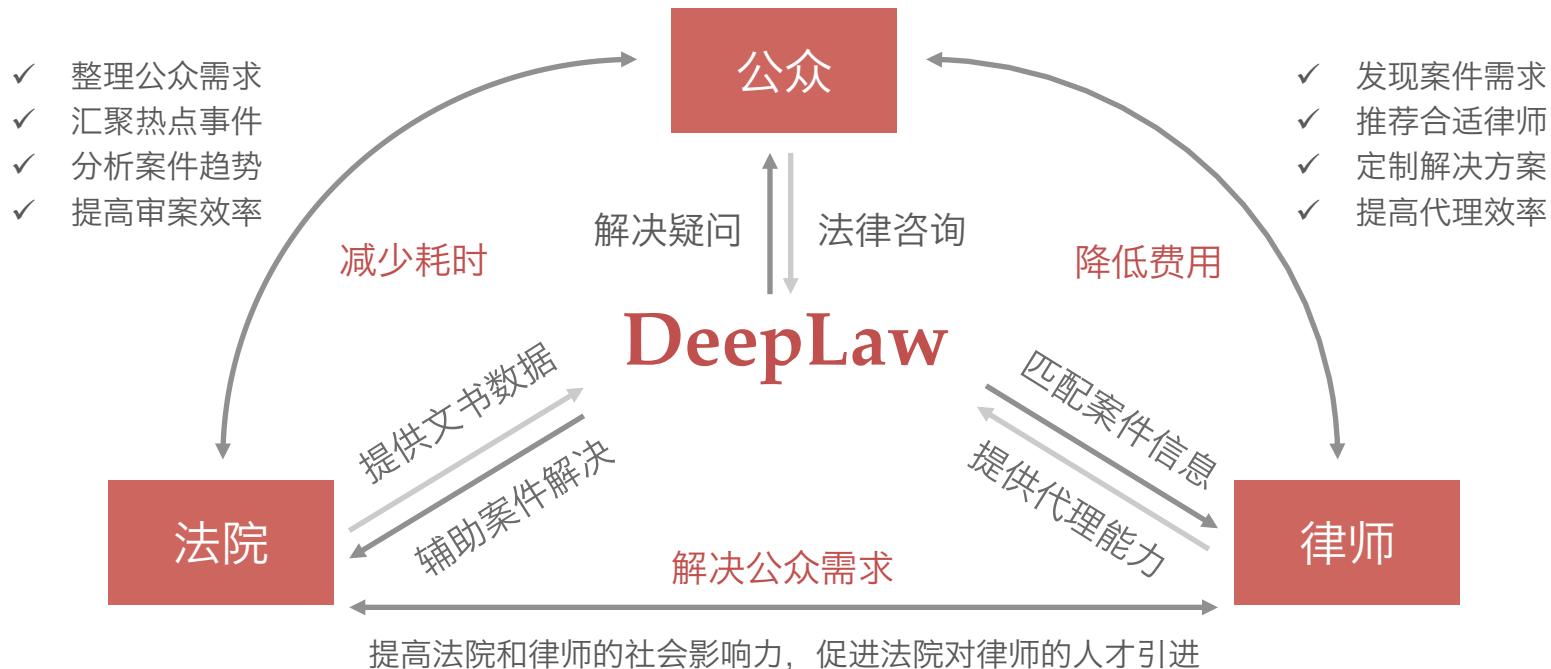
类自然语言和自由定制的
法律信息咨询
可谈心、可讨论、可交流



专业问答

基于法律法规和法律术语的
专业知识问答
可学习、可解惑、可答疑

商业模式与价值



理论框架

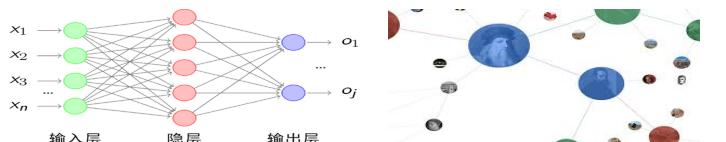
1 数据层

聊天语料 问答数据 领域本体 知识图谱



2 模型层

关键词搜索 模板匹配 检索评分 神经网络



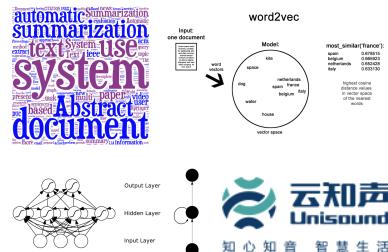
用户输入



对话输出

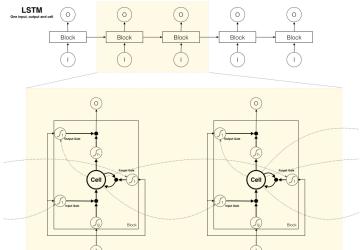
3 理解层

句法分析 语义理解
语境感知 语用计算



4 语言层

文本搜索 知识匹配
抽取式 生成式



数据层

语料和数据主要来自于 **互联网抓取**

聊天语料

基于 **新浪微博** 用户之间的 **评论互动** 获取 **聊天语料**



问答数据

基于 **法律咨询**、**华律网**、**法律快车** 等网站收集已解决的
咨询问题 和相应的 答案



领域本体

基于 **中国法院网** 采集专业 **法律术语**、**各类案由**所涉及的
本体 以及各类 **案件报道**

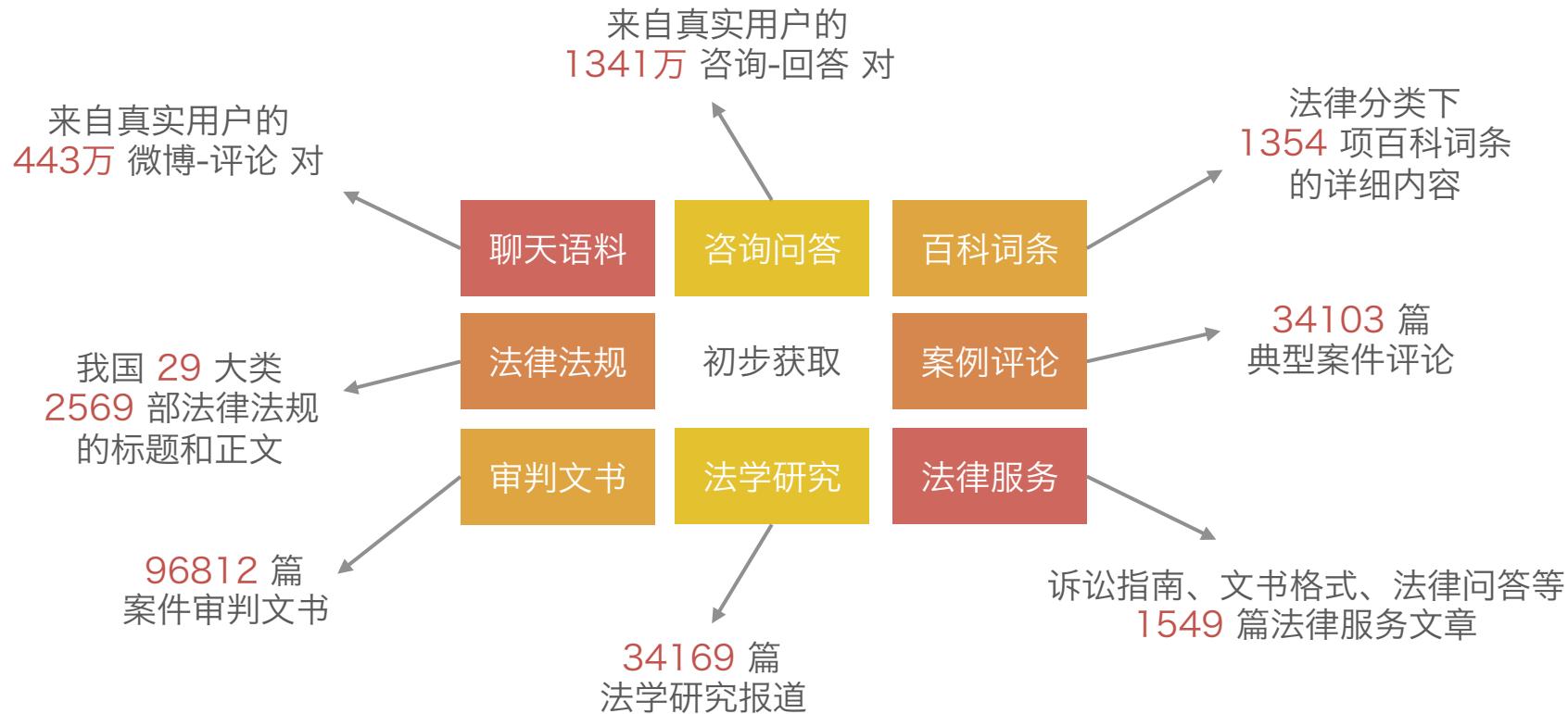


知识图谱

基于 **搜狗百科**、**百度百科**、**互动百科** 法律类别下的词条
构建法律垂直领域下的 **知识图谱**



数据资产



模型层

关键词搜索

提取输入中的关键词
并在语料库中搜索相
关的对话



< Post, Response >

检索评分

基于多特征检索评分
模型筛选最优对话



神经网络

基于LSTM等递归神
经网络模型对聊天数
据进行训练



模板匹配

基于预先定义好的文
本模版在语料库中匹
配相关的对话



关键词搜索



我想问一下
婚前同居合法吗



中文分词

TextRank
TF-IDF



婚前 同居 合法



Retrieval-based

直接 简单

但 效率低 效果差



模版匹配

-za-z0-9]+)([0-9]+)\.*\.

Rule-based

简单 准确

但 通用性差 整理成本高



你吃午饭了吗

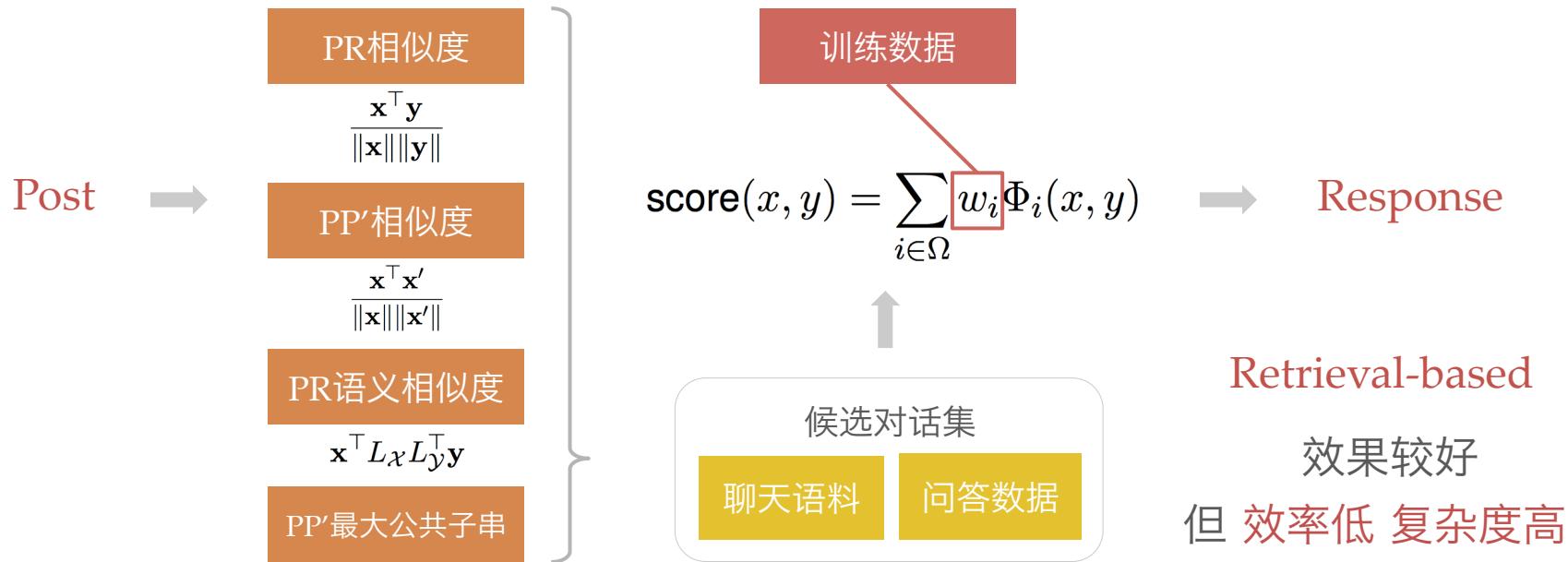


\S{0,5}吃\S{0,1}饭\S{0,5}



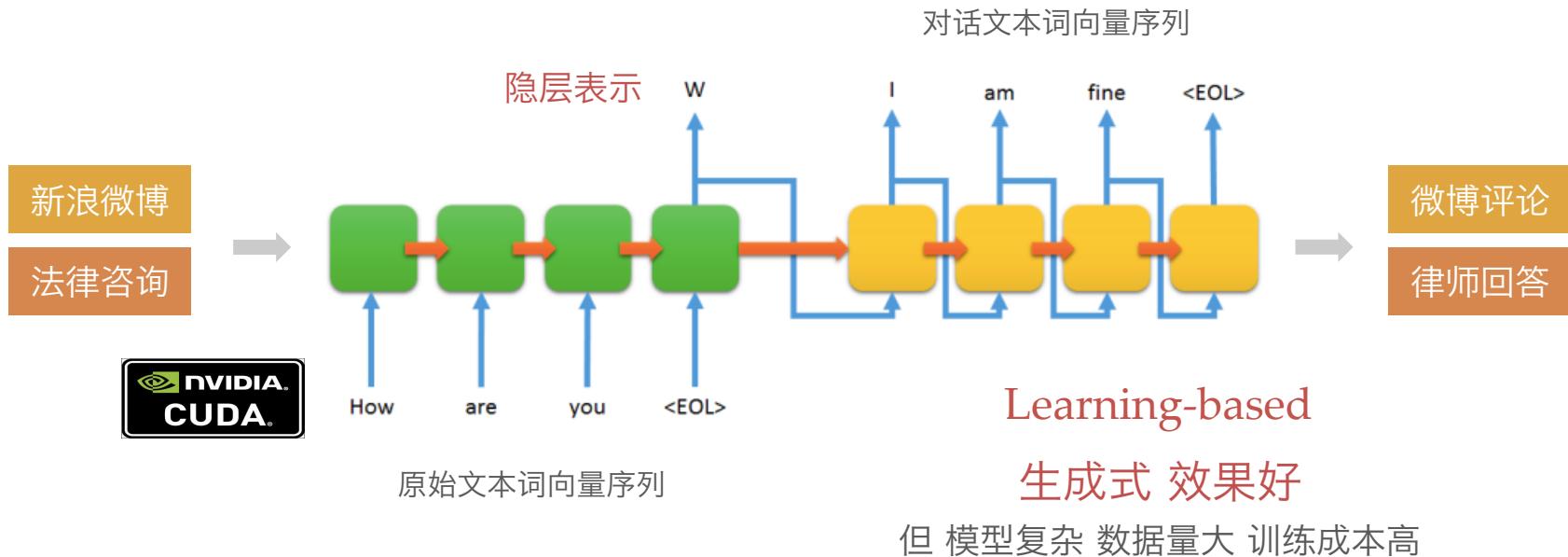
还没有，我好饿

检索评分

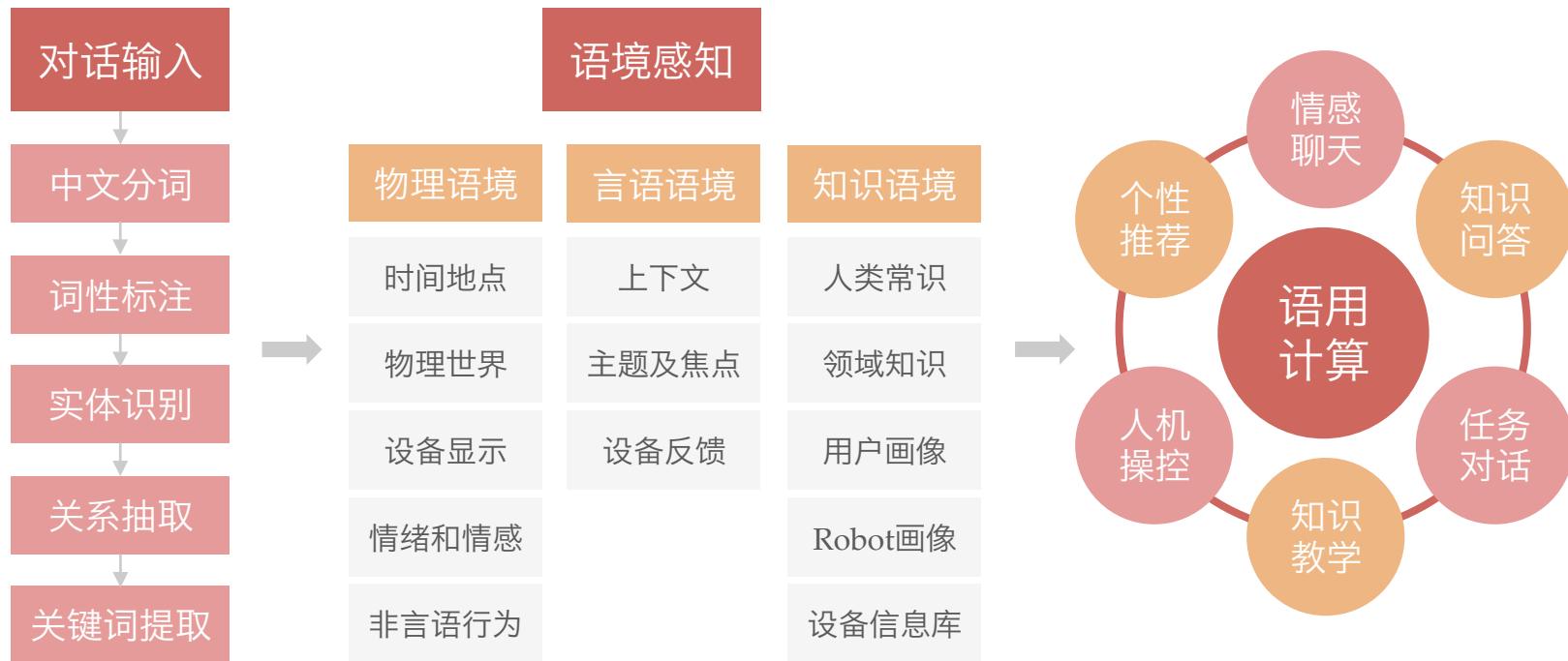


神经网络

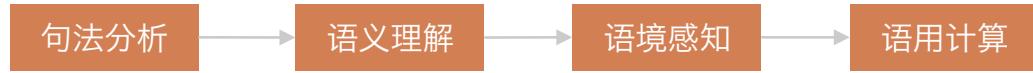
基于 深度LSTM网络 的 自编码器



理解层



语用计算



Post : 你好

Response : 上午好 ----- (聊天)

Post : 我的心情不好，遇上了一些事情

Response : 发生了什么事呢？我可以帮你什么吗？ ----- (情感)

Post : 我出了车祸，但是肇事司机不肯承担责任

Response : 按照法律规定，司机须按责任比例进行赔偿 ----- (问答)

Post : 那我应该怎么办呢

Response : 为您推荐道交案件律师，张某某 ----- (推荐)

Post : 有他的联系方式吗

Response : 手机号12233334444，邮箱zmm@gmail.com ----- (任务)

词嵌入



word2vec

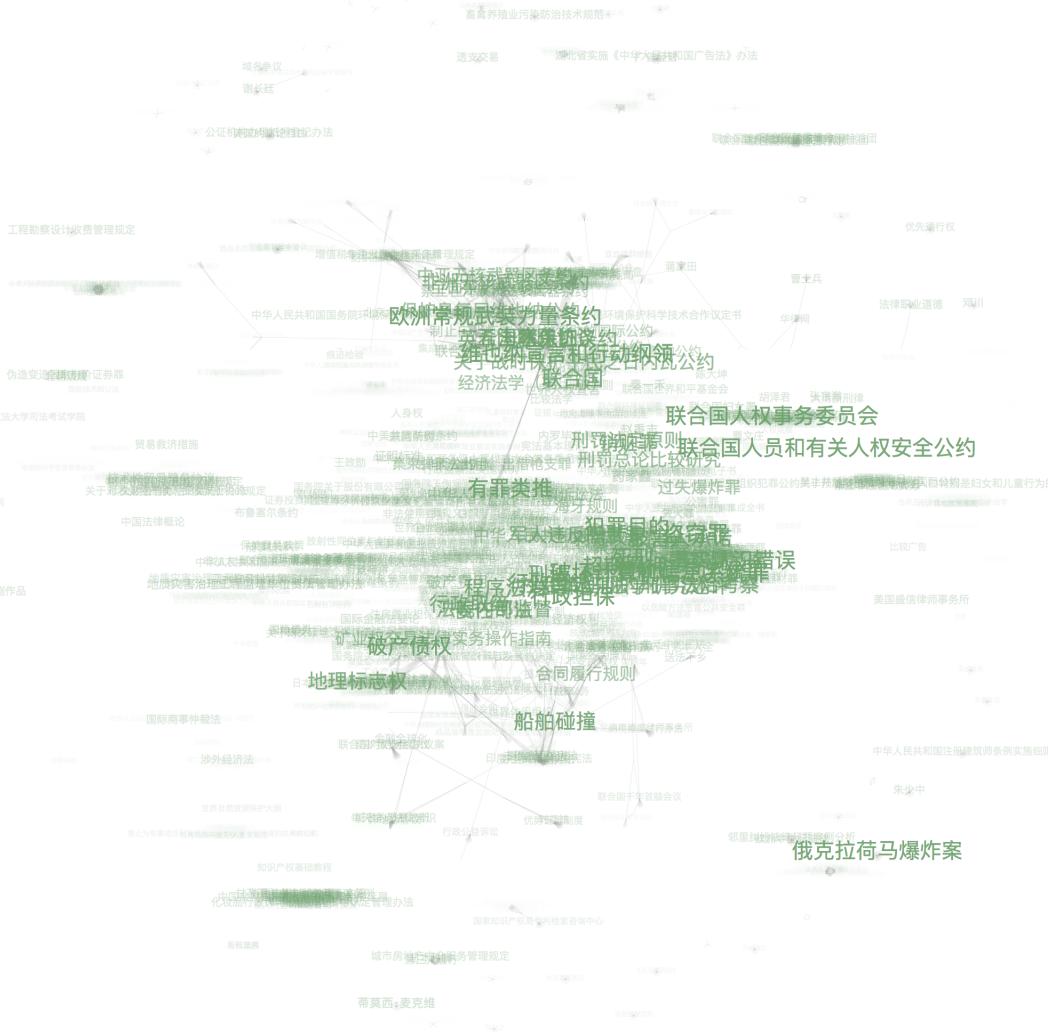
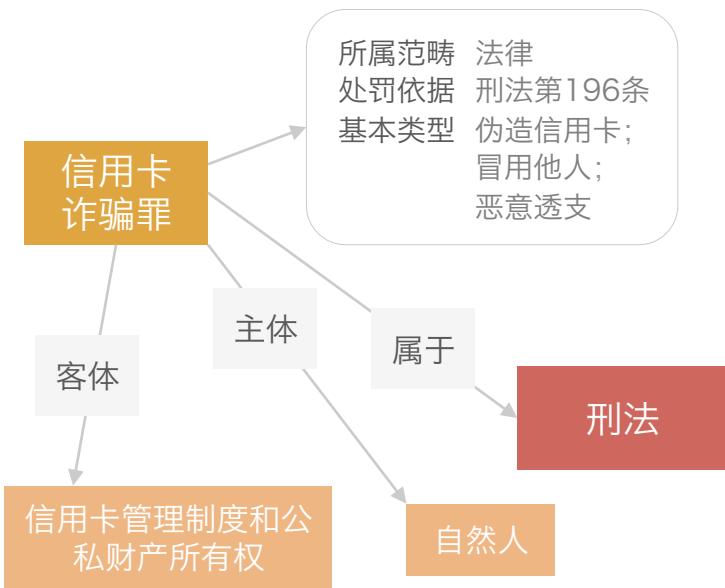


128 维词向量



知识图谱

基于 百科词条
抽取 知识 生成 图谱



语言层



技术实现

语言层

DeepLaw



Flask
web development,
one drop at a time

基于 Flask 实现产品交互界面

理解层

基于 Keras 实现深度学习
使用 Neo4j 存储知识图谱



farizrahman4u / seq2seq

模型层



jieba

word2vec

基于 Jieba 分词 进行初步文本处理
使用 Word2Vec 训练语言模型

数据层



基于 Scrapy 和 Redis 获取海量
多源异构数据并存储至 MySQL

应用

对话

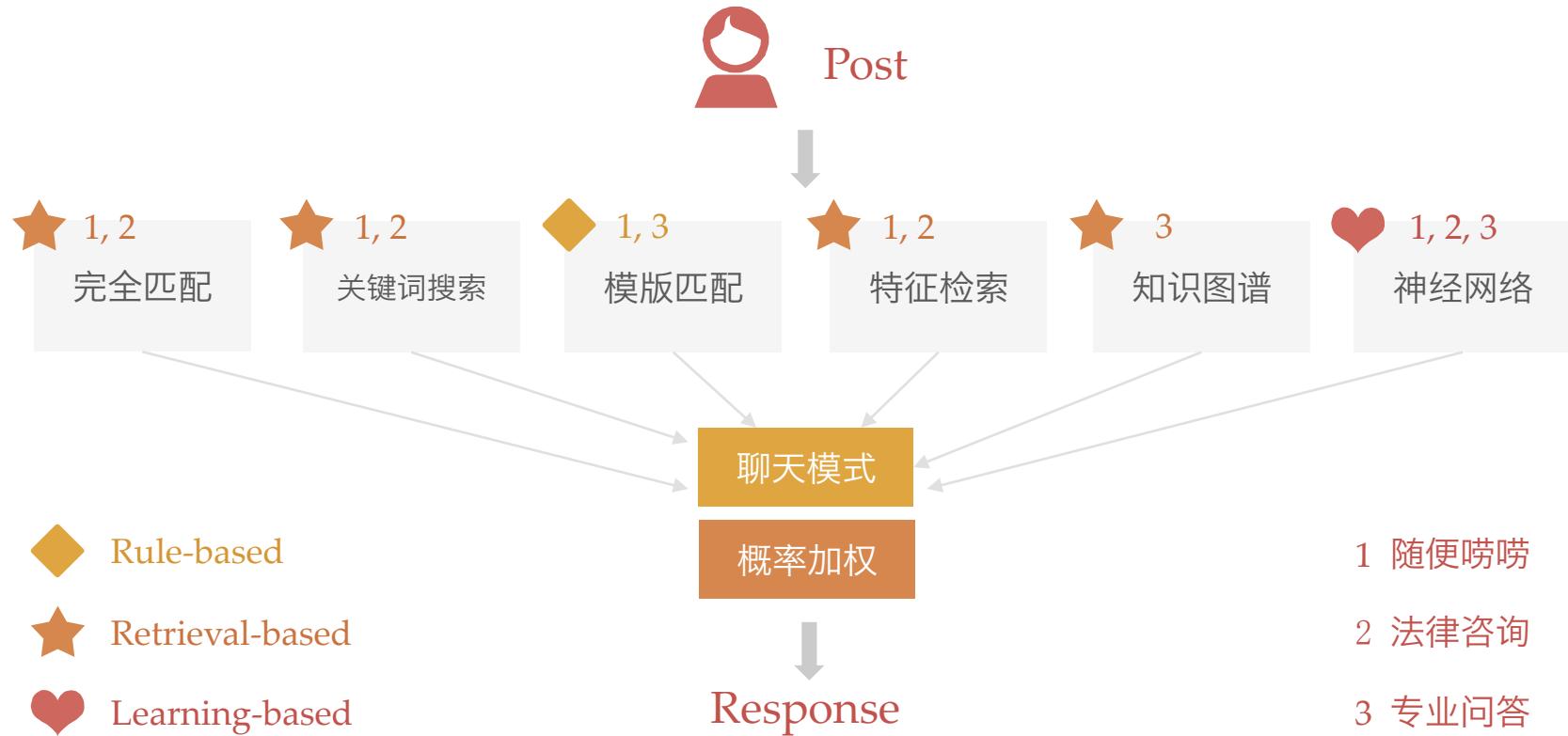
模型

数据

语料



技术流程



产品界面

DeepLaw 拟人互动界面

<http://zhanghonglun.cn/chatbot>

历史聊天记录



聊天模式选择

用户输入区

对话评估反馈

- ✓ 用户 **手动指定** 模式，随便唠唠、法律咨询或专业问答，DeepLaw 以相应模式进行聊天
- ✓ 用户 **直接输入** 聊天内容，DeepLaw 根据聊天语境 **自动判断** 最优模式，并进行聊天



DeepLaw

下一代人工智能法律顾问