

智能文档分析平台

一站式文档处理智能助手，释放文档新活力

• 目录

01 企业文档处理的痛点

02 智能文档分析平台架构概述

03 产品特点详解和实操

01

企业文档处理的痛点

企业文档类型繁多，处理场景多且压力大，急需智能化提效减负



- 01 企业文档“三多一少”：文档种类繁多、文档格式多样、版式布局多样、训练样本稀少
- 02 企业文档多为非固定版式，基于传统AI开发，定制化程度高，训练样本大、周期长，整体成本较高
- 03 基于传统AI开发的文档审查，规则配置复杂，缺乏逻辑推理判断能力

通用文档

合同文书

房屋租赁合同

出租方：_____ 职务：_____

承租方：_____ 职务：_____

根据有关法律的规定，甲乙双方经协商一致达成如下条款，以资遵守。

第一条：租赁物位置、面积、功能及用途。

1. 甲方将位于上海凤城路1820弄16号15号房屋（以下简称“房屋”），租赁给乙方使用。租赁物面积约1500平方米（最终以房屋产权证为准），用途为：_____。

第二条：租赁期限。

租赁期限为15年，即从2009年2月28日起至2024年3月27日止。

招投标书

工程施工招标

投标文件

招标编号：_____

投标文件内容：投标业商务标文件

投标人：_____

法定代表人或其委托代理人：_____

日期：_____年__月__日

保理登记证

中国人民银行征信中心

动产融资统一登记-初始登记

名称	浙江绿城房产有限公司	类型	企业
统一社会信用代码	913302033063089324	工商注册登记机关	浙江省市场监督管理局
法定代表人/负责人	张华	住所	浙江省宁波市鄞州区中河街道中河西路100号

法律判决书

广东省广州市中级人民法院

刑事裁定书

(2012)穗中法刑二终字第714号

被告人张某某，广东肇庆人，汉族，1975年7月28日出生，住广东省肇庆市高要区，因涉嫌非法吸收公众存款罪，于2012年4月11日被刑事拘留，同年5月18日被逮捕。现由肇庆市公安局移送审查起诉。

物流单据

EATON

国际物流单据

运单号	123456789
目的地	上海
重量	100kg

专业文档

广东省广州市中级人民法院

刑事裁定书

(2012)穗中法刑二终字第714号

被告人张某某，广东肇庆人，汉族，1975年7月28日出生，住广东省肇庆市高要区，因涉嫌非法吸收公众存款罪，于2012年4月11日被刑事拘留，同年5月18日被逮捕。现由肇庆市公安局移送审查起诉。

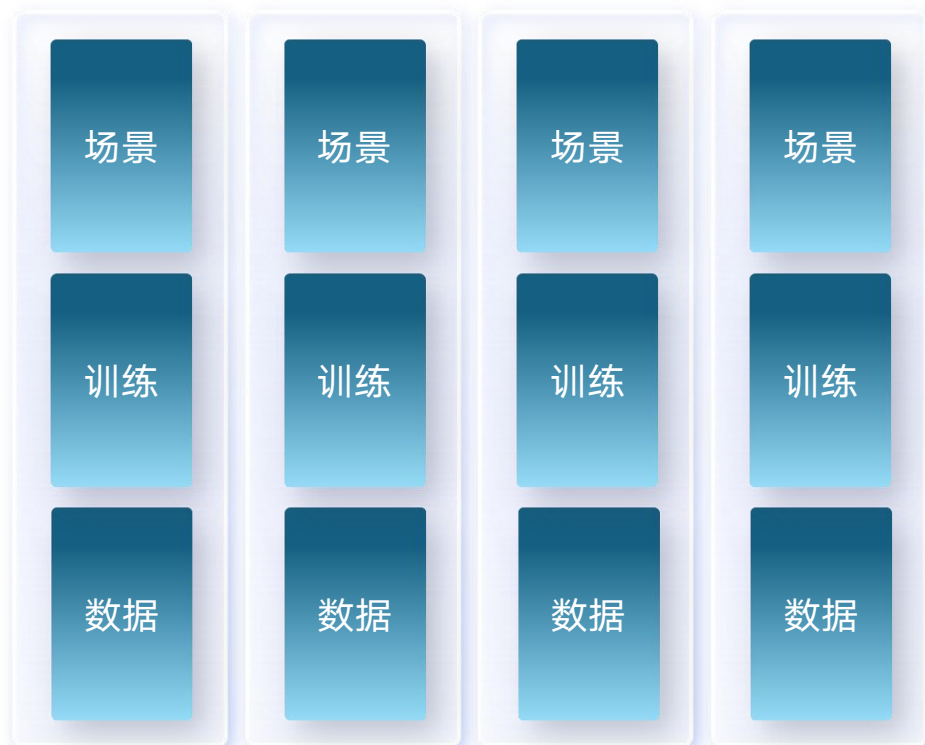
通用文档

专业文档

02 智能文档分析平台架构概述

语义大模型对文档智能化处理的革新

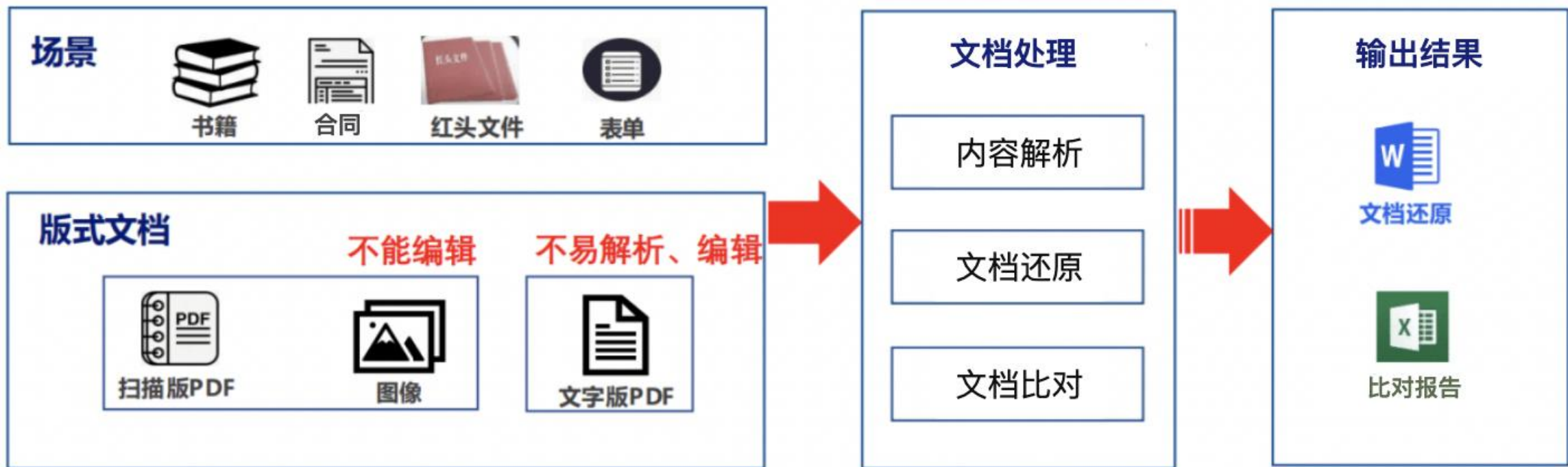
基于传统AI的应用开发



基于大模型的文档智能化分析



覆盖日常办公场景的高效工具



智能文档分析平台 TextMind



依托百度文心大模型、OCR等AI技术，为用户提供一站式文档智能化处理服务，帮助企业挖掘文档价值，提升处理效率，加强风险防控

行业实践

泛科技

能源

金融

工业

制造

零售

应用层

合同智能体

招采智能体

文档电子化

金融业务审查

检书业务审查

物流单证校验

能力层

文档格式转换

文档比对

合同审查

文档解析

文档抽取

标书审查

智能文档分析平台 跨模态文档大模型分析引擎

平台层

SaaS 服务（开箱即用）

预置审查 Prompt

Prompt 管理

预置场景模型

PaaS API+SDK（被集成）

数据标注

模型训练&推理

模型管理

多模态布局增强文档模型（ERNIE-Layout）

OCR识别结构化模型（VIMER-StrucText）

文档格式解析（DocParser）

文档版面分析与识别（OCR）

大模型基座层

ErnieBot 4.0 超千亿参数基础大模型

Ernie Speed 百亿参数基础大模型

基础设施层

异构算力

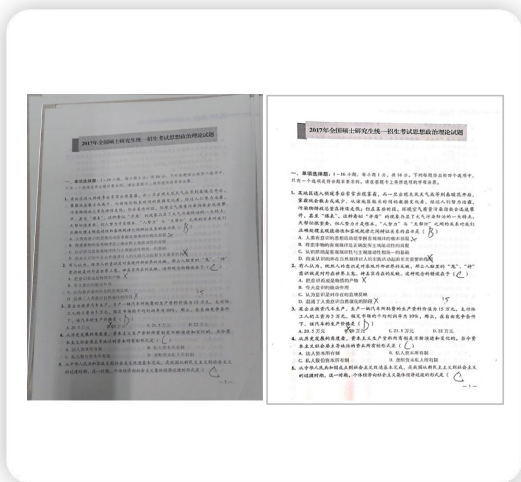
03 产品特点详解和实操

文档格式转换：多版式识别，多格式转换

- 文档格式转换平台是一款功能强大的多模态文档处理引擎，涵盖文档图像预处理、文档版式识别、文字识别、格式转换等全流程能力，为各类文档电子化处理场景（例如档案管理、信息录入、信息检索等）提供底层基础能力支撑



文档图像预处理：降低图像噪音干扰，提升文字识别效果



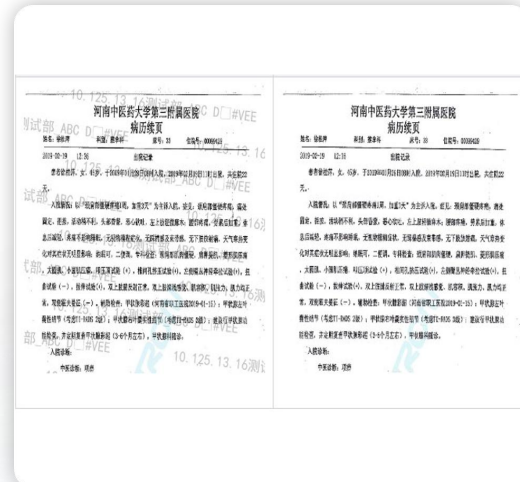
校正增强

提取主体内容并对其进行矫正，同时对图片进行增强，进一步提升图片清晰度



印章擦除

对原图中的印章进行擦除处理，减少因印章遮盖导致的文字识别错误



水印擦除

对原图中的水印进行擦除处理，减少因水印遮盖导致的文字识别错误



旋转检测与自动修正

可对文档方向进行检测和纠正，避免因旋转角度问题影响OCR识别

多模态文档识别引擎—对文档全要素进行识别解析（一）

针对办公类文档（如招标文件、合同、档案、行业报告、论文等）进行全要素识别解析。包括文档的版面元素、全文文字信息、表格结构信息、印章信息等。为各类文档电子化场景提供底层能力支撑。

- **文档版面识别：**对整份文档的版式结构和内容元素进行分析识别。**版式结构包括页眉、分栏、页脚、脚注和页码；内容元素包括配图、表格、标题、段落、目录、印章、公式等**，输出各板块的属性和位置。

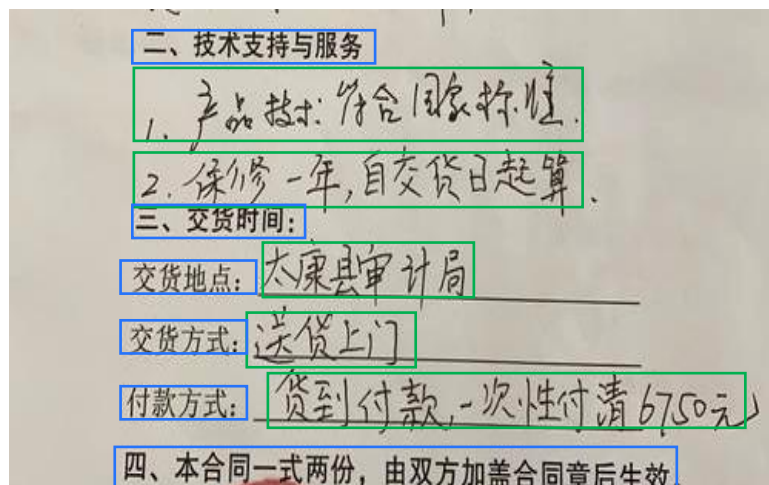


文档版面识别



多模态文档识别引擎—对文档全要素进行识别解析（二）

- **文字识别：**对全文文字内容和位置进行识别，支持中、英两种语言，可对手写、印刷体进行分类和识别
- **表格识别：**识别表格中单元格内容、行列结构和位置信息，支持有线表格、无线表格识别
- **印章识别：**检测并识别文档中的印章，输出文字内容、印章位置信息以及相关置信度，支持圆形章、椭圆形章、方形章等常见印章



印刷/手写文字识别

合并资产负债表

单位：元 币种：人民币

项目	2018年12月31日	2019年1月1日	调整数
流动资产：			
货币资金	1,885,297,747.60	1,885,297,747.60	
衍生金融资产			
应收票据			
应收账款	7,871,933.84	7,871,933.84	
应收款项融资			
预付款项	85,234,762.60	85,234,762.60	
其他应收款	636,498,971.19	636,498,971.19	
其中：应收利息			
应收股利			
存货	8,530,771,284.66	8,530,771,284.66	
持有待售资产	43,805,119.64	43,805,119.64	
一年内到期的非流动资产			
其他流动资产	358,247,448.31	358,247,448.31	
流动资产合计	11,497,727,267.84	11,497,727,267.84	

表格结构识别

JG案例

应用于某JG服务商的可视化文档管理系统，对图片/PDF文件进行全文识别并还原为word格式的文档，进行存档和信息摘录管理，同时对外文文档进行翻译并还原布局

公检法案例

应用于某省检察院的文档管理系统，将图片/PDF文件转换为Word/Excel或双层PDF文件进行存档，方便复制信息和文字检索

电网案例

应用于某省电力公司内部办公系统软件，为内部员工提供可视化平台工具，将图片/PDF文件转换为Word/Excel文件，满足日常办公需要。同时提供api接口，为企业办公系统ocr赋能提供算法支撑

金融案例

应用于金融文档处理平台，对图片/PDF文件进行识别，利用文档版式分析结果和文字识别结果，抽取产业知识，构建知识图谱，进行金融数据预测，辅助智能写作

文档比对：精准发现版本间差异，防篡改

- 比对文档之间的内容信息，**精准识别出文档内容的差异点**，分钟级自动列出文档差异列表，助力合同审核、防篡改等业务场景提效

业务处理效率
提高75%

业务流程准确率
提高90%

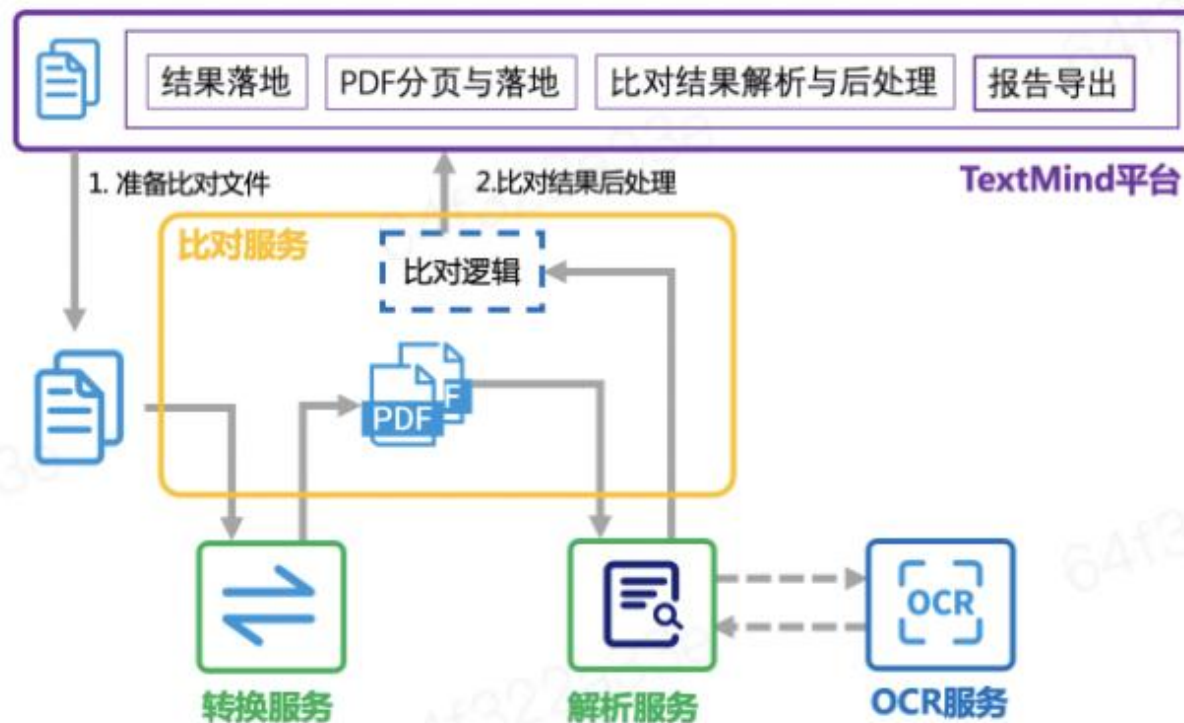


比对结果精准



可视化界面

- 识别分块/页眉页脚等布局
- 支持表格中的内容比对
- 支持旋转、翻转等实际样式变化
- 识别水印、盖章等图层信息
- 支持长文档跨页比对
- 私有化支持1vs1、1vsN、特定页比对
- 完整的前端界面，一键式操作
- 结果简单清晰，差异点快速定位
- 可人工审核修正比对结果



将输入文档转换成PDF格式

如果是版式或者不可编辑扫描件，则需要调用OCR服务

工业案例

应用于某车企的内部办公系统软件，合同签署前的电子版Word和扫描件PDF比对，防止合同出现篡改、错漏、签字版与电子版不一致的情况

能源案例

应用于某能源企业的合同管理系统，合同管理系统中内置比对工具，合同起草-审批过程中，多次修改产生多版本文件，确保文件正确

金融案例

应用于某省银行公司内部办公系统软件，比较不同授权委托书之间的文字和印章差异，验证委托书的真实性和一致性

合同审查：大模型合规性审查，加强合同风险防控

文心大模型驱动，支持20+种合同场景审查，提供专业的风险评估及建议，综合审查准确率达85%+，效果名列前茅

智能化

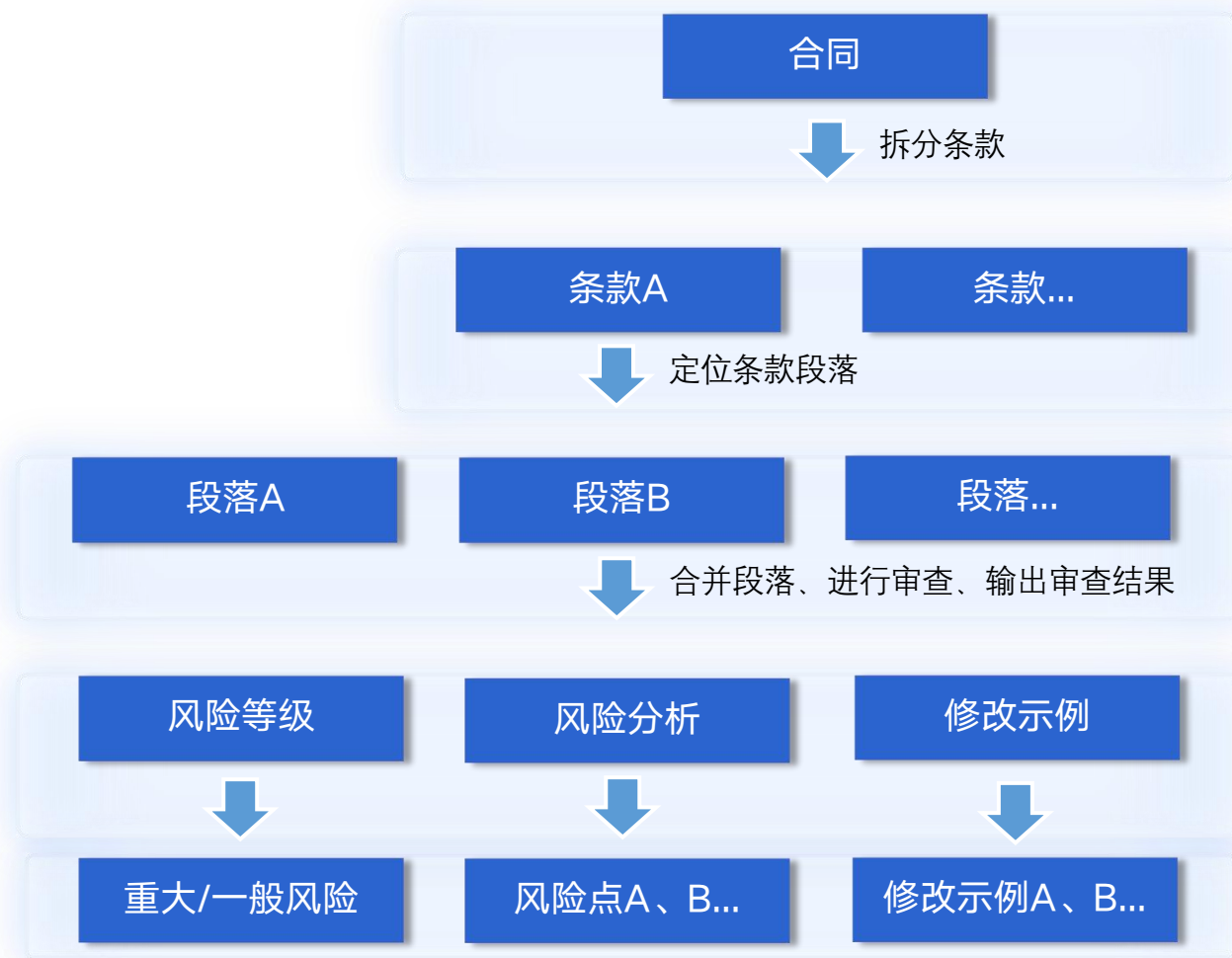
理解复杂法律文本，自动定位和分析关键条款

多样化

审查模型可灵活配置
高效满足定制化需求

持续进化

持续学习法律法规
保持知识最新性



合同审查：自定义清单配置

为满足不同企业的独特业务需求，用户可灵活自定义审查清单，包括条款的增添、删除与修改，以实现个性化审查标准

The screenshot displays the '自定义清单配置' (Custom Review List Configuration) interface. On the left, a sidebar contains a '新建审查清单' (New Review List) button and a '清单列表' (List of Lists) section with '我的清单' (My Lists) and '预置清单' (Predefined Lists). The main area is titled '我是一个自定义清单' (I am a Custom Review List) and shows configuration for a specific list. It includes a '条款列表' (List of Clauses) with checkboxes for various clause types like '主体信息条款' (Party Information Clause), '价款与结算条款' (Price and Settlement Clause), etc. The right side shows the configuration for two selected clauses: '主体信息条款' and '价款与结算条款'. Each clause has a text field for its description and a '风险点' (Risk Points) section with checkboxes for specific risks. A modal dialog is open, asking '确定删除该条款下的所有内容吗?' (Are you sure you want to delete all content under this clause?).

新建审查清单

清单列表

我的清单

我是一个自定义清单

我是一个自定义清单2

我是一个自定义清单3

我是一个自定义清单4

预置清单

买卖合同

租赁合同

技术开发合同

劳动合同

租赁合同

技术开发合同

劳动合同

我是一个自定义清单 展示填写示例 保存 删除

清单来源: 买卖合同-买方 | 清单ID: am-fpg4sq90itdk | 清单描述: 我是一个清单描述

条款列表 + 添加

主体信息条款

价款与结算条款

交付条款

验收条款

质量与质量保证条款

标的物风险转移条款

违约责任条款

知识产权条款

合同生效与有效期条款

合同变更条款

通知条款

主体信息条款

*条款说明: 主体信息条款是指合同首页或落款签署页明确合同当事人的内容, 需要明确签约主体名称和地址以及联系人信息。

*风险点: 主体信息缺失 主体信息不一致

+ 添加自定义风险点

价款与结算条款

*条款说明: 价款与结算条款包括货物价款、价款单位和税率、价款的支付方式、价款支付的时间或支付条件、发票开具、银行账号等必备内

*风险点: 缺失标的物价款信息 缺失价款支付信息 缺失收款方信息 交付前支付全款

*名称: 我是一个风险点名称

*说明: 我是一个风险说明

+ 添加自定义风险点

确定删除该条款下的所有内容吗?

取消 确定

合同审查：数据标注+精调

用户可上传并标注数据，并能将标注数据与审查清单相结合，生成高质量的可用于模型精调的数据



精调数据集名称	精调数据集类型	最新版本	版本描述	导入状态	数据量	发布状态	更新时间	操作
精调数据集-20241227-验证	召回	V1	-	● 已导入	4765	● 已发布	2024-12-27 19:37:34	导入 发布 去精调 更多
精调数据集-20241227-训...	召回	V1	-	● 已导入	201	● 已发布	2024-12-27 19:37:32	导入 发布 去精调 更多
精调数据集-20241121-174...	审查	V1	1	● 已导入	1	● 未发布	2024-11-21 17:44:30	导入 发布 去精调 更多

合同审查：大模型 vs 小模型

	大模型方案	小模型方案	备注
审查维度	<ul style="list-style-type: none">■ 形式与实质审查兼具✓ 在一致性、完整性等形审之上，通过深度语义理解对合同全文进行合规性审查，确保符合法律法规和商业伦理标准✓ 实现从律师/法务视角的形式、合规性和有效性的全方位审查	<ul style="list-style-type: none">■ 仅支持形式审查✓ 关注合同文件本身的结构一致性和完整性，辅助完成合同初审的校对，如金额大小写一致性、价格与数量匹配性等	<ul style="list-style-type: none">■ 审查条款举例--“乙方违约责任条款”：<ul style="list-style-type: none">• 原文：“如违反上述情形，违约方应赔偿业务合作总金额的50%的赔偿责任。在任何情况下，乙方基于本协议及订单的赔偿责任总额，不超过甲方已支付的服务费用；若损失涉及的服务期限超过12个月，则乙方最大赔偿责任不超过损害前6个月的费用。”
泛化性	<ul style="list-style-type: none">■ 高度灵活✓ 仅需少量或无需额外数据，基于大模型SFT与Prompt工程即可精准识别不同合同场景的法律风险✓ 高效拓展场景和条款，准备和部署期更短	<ul style="list-style-type: none">■ 场景依赖性强✓ 每个合同场景均需经历数据获取、标注、模型训练，泛化能力受限✓ 拓展场景或条款的前期准备期耗时长、成本高	<ul style="list-style-type: none">• 小模型审查结果：违约金比例过高，超过合同总金额的50%，建议修改比例。• 大模型审查结果：1) 违约金比例过高，超过合同总金额的50%，建议修改比例。2) 该条款限制了乙方的赔偿责任，可能导致甲方损失无法完全得到赔偿，建议删除乙方赔偿责任的限制或增加例外情形，以保障甲方权益。

合同智能审查案例1——某能源客户私有化部署“合同智能化应用”



项目背景

- 某能源集团提出：聚焦合同风险防控，从严加强合同基础管理
- 集团年处理量70多万份合同数据，其中7万份为非标合同数据，审核周期长、压力大
- 国内法务审核领域人员流动率较高，服务团队持续保障高质量审核挑战大

效果价值

- 构建6个合同场景模型，涉及300+个合同风险点，覆盖合同审查全部要点，完善了合同审查体系
- 缩短审核周期，合同审批效率提升68%
- 人工智能审核水平相当于工作非常细致的初级律师，助力法务审核团队始终保持高水准

解决方案

共建领先的油气领域合同智能化应用平台

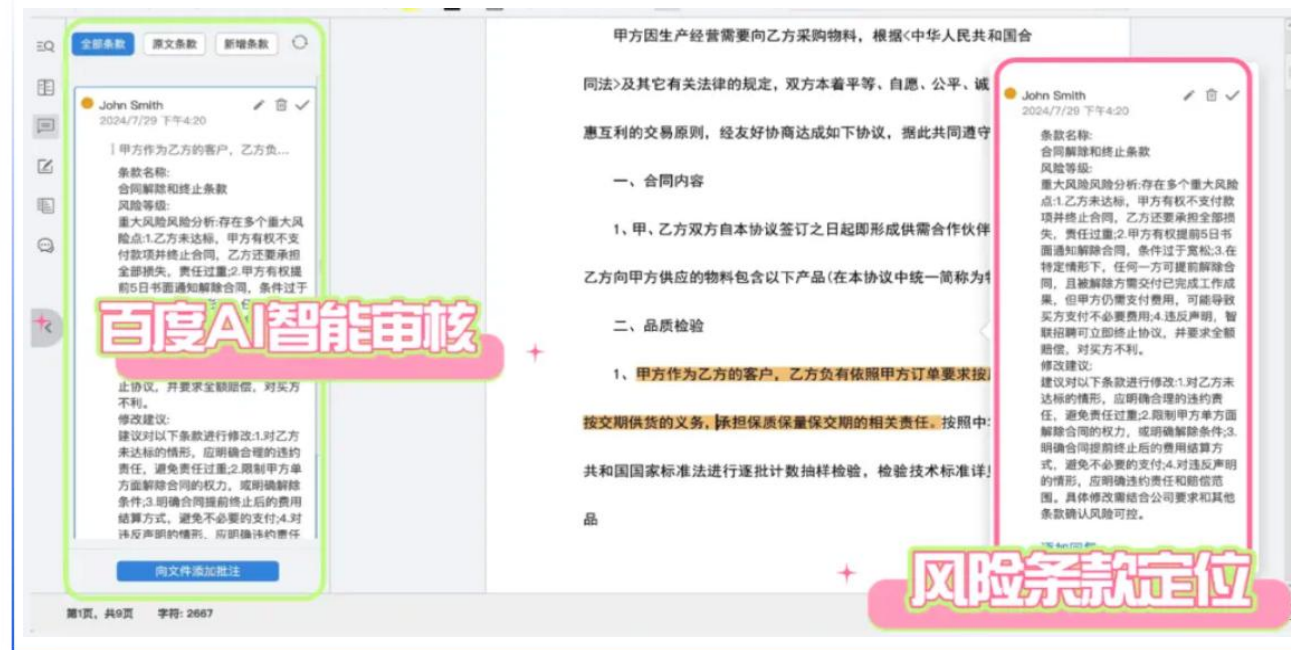


合同智能审查案例2——某泛科技公有云“合同智能审查”



- **解决方案：**基于文心大模型4.0的合同审查，围绕客户法审体系构建审查条款库，为企业提供量身定制的实质审查AI服务，满足客户合规性审查需求，大幅提升合同审查效率；同时提供的在线编辑与一键清稿功能，助力客户提升在合同修改与保存环节的处理效率

- **实质审查：**精准识别潜在的合规性、一致性及完整性等风险，为法务人员提供清晰、精准的审核结果
- **条款定制：**基于客户自身法审规则体系，构建审查条款库
- **在线编辑：**根据审核结果一键完成合同修改和保存，保障合同全流程在线不落地处理
- **一键清稿：**快速清理修订痕迹，生成合同最终版本



■ 效果价值：

- AI审核准确率可达85-90%，助力法务审核团队始终保持高水准
- 缩短了审核周期，大幅合同审批效率
- 为公司AI技术战略树立标杆示范

Thanks