**HyperEventGraph 演示文稿 (最终版)**

**主题风格:** 严谨、专业、清晰。采用白色/浅灰色背景，以上海大学Logo和标准色（蓝/红）作为点缀，确保学术报告的正式感和专业性。

**幻灯片逐页详解**

**Slide 1: 标题页**

* **页眉/页脚:** 左上角放置校徽，右下角放置报告时间。
* **页面内容:**
  + **大标题:** 超图与大语言模型：风险分析思维链的构建与应用
  + **副标题 (可选):** HyperEventGraph 系统设计与实现
  + **汇报人:** [汇报人姓名] (学术型硕士)
  + **学号:** [学号]
  + **时间:** 2025年8月1日
* **演讲者备注:**
  + 开场白，介绍自己和报告主题。明确指出本次报告将围绕一个名为HyperEventGraph的系统，探讨如何利用超图和大语言模型技术，解决金融风控等领域的复杂问题。

**Slide 2: 议程 [新增/强化页]**

* **标题:** 报告大纲 (Agenda)
* **核心信息:** 为听众提供清晰的报告结构和预期。
* **视觉设计:**
  + 使用简洁的垂直时间线或编号列表，每个条目旁配有对应的图标。
  + **内容:**
    1. **研究背景与动机 (The Challenge):** [一个放大镜图标] 我们面临什么问题？
    2. **核心问题与框架 (The Framework):** [一个流程图图标] 为何选择超图及我们的解决思路。
    3. **系统设计与实现 (The System):** [一个齿轮或架构图图标] 系统架构与关键技术细节。
    4. **核心应用与展望 (The Payoff):** [一个灯泡或大脑图标] 系统能带来什么价值以及未来工作。
* **演讲者备注:**
  + 快速过一遍议程，让听众对报告的整体逻辑有一个清晰的认识，知道接下来会讲什么。

**Slide 3: 研究背景：行业痛点与系统目标**

* **标题:** 研究背景：从行业痛点到系统目标
* **页面布局:** 左右分栏，布局清晰，视觉引导性强。
* **左栏：行业三大痛点 [cite: 17-29]**
  + **标题:** 传统风险分析的困境
  + **内容:**
    1. **财务指标滞后:** 图标为一个缓慢变化甚至下跌的股价K线图。文字说明：“无法及时捕捉‘事件→事理→风险’的早期信号，导致风险识别延后。”
    2. **信息孤岛严重:** 图标为三个分离的文档（新闻、招股书、社交媒体），中间有隔断墙。文字说明：“多源信息割裂，难以形成统一的风险认知。”
    3. **自动化工具缺乏:** 图标为一个分析师在堆积如山的文件前感到困惑。文字说明：“银行信审、投资人等关键角色缺少高效、自动化的风险洞察与预警工具。”
* **右栏：系统两大核心目标 [cite: 30-40]**
  + **标题:** 本研究旨在实现
  + **内容:**
    1. **自动串联多源信息:**
       - **核心流程图:** 设计一个清晰的、从左到右的流程图：[一个新闻图标] → [一个超图结构图标] → [一个链条图标]。下方分别标注“领域新闻”、“事理超图”、“风险链”。
       - 文字说明：“实现‘领域新闻→事理超图→风险链’的自动化构建与联通。”
    2. **实时业务动向预测:**
       - 在“风险链”方框下方，列出系统可输出的关键价值点：
         * 下一步可能动作
         * 潜在连锁风险
         * 可信解释路径
* **演讲者备注:**
  + 从行业普遍面临的挑战入手，强调传统方法的局限性，引出本研究的必要性。
  + 清晰地阐述系统希望达成的两个核心目标，特别是“领域新闻→事理超图→风险链”这一核心理念，为后续内容铺垫。

**Slide 4: 核心问题：为何必须是超图？**

* **标题:** 核心问题定义：传统图谱的表示瓶颈
* **核心信息:** 直观对比传统知识图谱与超关系知识图谱在表达复杂n元关系时的根本差异。
* **页面布局:** 左右并排对比图。
* **左侧：传统知识图谱 (信息拆分) [cite: 74-88]**
  + **标题:** 传统KG表示（二元关系）
  + **图示:** 以“公司并购”为例，一个“收购方”节点通过**三条独立的边**分别指向“被收购方”、“交易金额”、“生效日期”三个节点。
  + **标注:** 在图下方用红色文字强调：“信息被割裂，无法表达完整事件。面对n元事实，必须拆解为多个三元组，导致高阶交互的整体性语义丢失。”
* **右侧：超关系知识图谱 (完整表达) [cite: 89-103]**
  + **标题:** 超关系KG表示（n元关系）
  + **图示:** 同样以“公司并购”为例，用\*\*一个大的、半透明的“超边”\*\*同时框住并连接“收购方”、“被收购方”、“交易金额”、“生效日期”四个节点。
  + **标注:** 在图下方用绿色文字强调：“直接、完整地表示实体间的高阶交互关系，保持事件的内在结构和全部信息。”
* **演讲者备注:**
  + 这是整个报告的技术立足点，必须讲清楚。
  + 通过强烈的视觉对比，让听众瞬间理解为什么需要引入“超图”这一概念，以及它解决了什么根本问题。

**Slide 5: 研究框架与思路 [新增/强化页]**

* **标题:** 研究框架与核心思路
* **核心信息:** 宏观上展示本研究从构建、精炼到推理的完整框架，并映射对应的研究思路。
* **页面布局:** 上下结构。
* **上部：三阶段核心框架 [cite: 52-64]**
  + **流程图:** 一个从左到右的、包含三个主要阶段的流程图。
    1. **阶段一：事理图谱构建:** 输入是“非结构化文本”，输出是“原始事理图谱”。
    2. **阶段二：知识精炼:** 输入是“原始事理图谱”，通过“内部超关系补全”和“事实纯化与验证”，输出是“精炼事理图谱”。
    3. **阶段三：高级推理:** 输入是“精炼事理图谱”，输出是“高级推理能力”（如多跳问答、因果推断）。
* **下部：研究思路 (挑战 → 方法) [cite: 111-125]**
  + **三栏对应关系图:**
    1. **挑战 (难点问题):**
       1. 缺乏领域先验知识
       2. 自动构建存在噪声
       3. 复杂问题解答的可解释性
    2. **方法 (研究内容):**
       1. 基于大模型的图谱构建方法
       2. 基于扩散模型的超关系图纯化方法
       3. 面向金融风控的风险思维链构造方法
    3. 用箭头将对应的“挑战”和“方法”连接起来。
* **演讲者备注:**
  + 上部分讲“做什么”，展示整个研究的宏观流程。
  + 下部分讲“怎么做”，将具体的研究方法与要解决的挑战一一对应，体现研究思路的逻辑性和完整性。

**Slide 6: 系统架构图**

* **标题:** HyperEventGraph 系统架构
* **核心信息:** 展示系统从数据输入到知识存储的五层架构，体现工程实现的严谨性。
* **视觉设计:**
  + **分层架构图:** 完全复现并美化PDF第8页的五层架构图 [cite: 333-358]。
    - **Layer 1: 数据输入层** (原始文本数据, 结构化数据)
    - **Layer 2: 数据处理层** (事件抽取引擎, 实体标准化, 人工审核)
    - **Layer 3: Cortex 智能上下文引擎** (算法粗聚类, 簇内去噪, 故事单元生成)
    - **Layer 4: 关系分析层** (事件关系识别, 逻辑关系推断, 知识融合)
    - **Layer 5: 知识存储层** (Neo4j 图数据库 & ChromaDB 向量数据库)
  + **动画与标注:**
    - 使用自下而上的箭头动画，表示数据的流动方向。
    - 在每一层的右侧，用简洁的文字标注其核心职责，如“将非结构化的原始文本数据转换为规范化的结构化数据”等。

**Slide 7: 项目开发路线图 (Gantt Chart)**

* **标题:** 项目开发路线图 (V3.1 → V4.0)
* **核心信息:** 以甘特图形式展示项目从基础链路验证到知识推理引擎升级的完整开发过程，体现项目管理的规划性。
* **视觉设计:**
  + **甘特图:** 复现并美化PDF第14页的甘特图 [cite: 480-507]。
    - **横轴:** 时间轴 (01-01 至 09-01)。
    - **纵轴:** 任务列表。
    - **两个主要阶段（用背景色块区分）:**
      1. **V3.1 (02-01 ~ 04-01): 核心目标 - 搭建AI+人工协同的初步链路。** 包含任务1-3。
      2. **V4.0 (04-01 ~ 09-01): 核心目标 - 升级为知识推理引擎。** 包含任务13-21。
    - **关键里程碑:** 在 08-01 时间点用一条红色竖线标注“当前进度”。
* **演讲者备注:**
  + 简要说明项目是如何通过两个大的版本迭代，逐步实现最终目标的。
  + 强调V3.1为V4.0提供了基础的结构化数据，体现了迭代开发的思想。

**Slide 8: 实现细节(I): 事件Schema设计**

* **标题:** 实现细节(I): 事件Schema - 领域知识的形式化
* **核心信息:** 展示系统如何将特定领域的事件知识，形式化为机器可严格处理的结构化定义。
* **页面布局:** 左右分栏。
* **左栏：领域事件类型定义 [cite: 361-377]**
  + **表格:** 精简展示PDF第9页的事件类型表格，突出“金融领域”和“集成电路领域”的几个典型事件。
    - **金融领域:** 公司并购、高管变动
    - **集成电路领域:** 供应链风险、新产品发布
* **右栏：具体Schema示例 - “合作合资事件” [cite: 378-401]**
  + **代码框:** 以代码高亮的形式，完整展示“合作合资事件”的JSON Schema。
  + **标注:** 用箭头和注释，对Schema中的关键部分进行说明，如：
    - "description": “对事件的自然语言描述”
    - "properties": “定义事件的核心要素（n元关系）”
    - "required": “定义构成该事件的必要元素”
* **演讲者备注:**
  + 强调Schema是后续所有自动化处理的基础和“法律”。
  + 通过具体示例，让听众理解系统是如何将抽象的领域概念，转化为精确的、机器可读的指令的。

**Slide 9: 实现细节(II): LLM驱动的事件抽取**

* **标题:** 实现细节(II): 基于提示词工程的事件抽取
* **核心信息:** 展示如何通过精巧的提示词工程，引导大语言模型（Deepseek R1）进行精准、事实驱动的n元关系抽取。
* **页面布局:** 上下结构。
* **上部：核心提示词指令 [cite: 405-424]**
  + **代码框:** 截取PDF第10页的关键提示词片段。
  + **高亮:** 用不同颜色的高亮框出几个核心指令：
    - **角色定义:** 你是一名专业的信息抽取系统...
    - **事实约束:** 抽取一个或多个已发生的事实事件
    - **绝对禁止项:** > 绝对禁止抽取预测类、观点类、推测类内容为事件。 (用红色框突出)
* **下部：抽取示例 (输入 → 输出) [cite: 428-448]**
  + **输入文本框:** 显示包含事实与预测的混合新闻原文。
  + **输出代码框:** 显示LLM返回的、只包含事实部分的、结构化的JSON事件数组。
  + **动画:** 用一个从“输入”到“输出”的箭头，并在箭头上标注“LLM (Deepseek R1)”，表示处理过程。
* **演讲者备注:**
  + 强调提示词工程在本研究中的重要性，特别是如何通过指令约束保证抽取结果的“高保真度”。
  + 通过输入输出的对比，生动展示了系统是如何从嘈杂的文本中精准提取价值信息的。

**Slide 10: 实现细节(III): Cortex引擎与知识精炼 [新增/强化页]**

* **标题:** 实现细节(III): Cortex引擎 - 上下文重建与知识精炼
* **核心信息:** 详细阐述Cortex模块如何通过“算法粗聚类 + LLM精炼”的两阶段策略，将离散事件组织成有上下文的“故事单元”。
* **视觉设计:**
  + **详细流程图 (基于PDF第12页 [cite: 449-468]):**
    1. **输入:** 一堆状态为 pending\_clustering 的“离散事件”图标。
    2. **阶段一：算法粗聚类:**
       - 一个标有“DBSCAN”的模块接收事件。
       - 下方用公式标注其核心：混合距离 = w \* 语义距离 + (1-w) \* 实体距离
       - 输出是几个被虚线圈起来的“粗簇”。
    3. **阶段二：LLM精炼 (分支处理):**
       - **小簇路径:** 直接进入一个LLM图标，输出“故事单元”。
       - **大簇路径:** 进入一个更复杂的子流程，包含三个步骤：“摘要 → 检索 → 扩展”，然后再进入LLM图标，输出“故事单元”。
    4. **最终输出:** 所有“故事单元”的状态被更新为 pending\_relationship\_analysis。
* **演讲者备注:**
  + 这是本研究的核心创新点之一，需要详细讲解。
  + 解释为什么需要这个模块：为了解决单个事件信息量不足，难以进行高质量关系分析的问题。
  + 强调“混合距离”和针对大簇的“摘要-检索-扩展”策略是其中的关键技术细节。

**Slide 11: 实现细节(IV): 知识混合搜索 [新增/强化页]**

* **标题:** 实现细节(IV): 知识混合搜索 - 赋能高级推理
* **核心信息:** 展示系统如何结合图数据库（精确连接）和向量数据库（语义相似），为高级推理提供全面、立体的上下文信息。
* **视觉设计:**
  + **UI界面模拟 (基于PDF第13页 [cite: 469-479]):**
    - **顶部:** 一个搜索框，用户输入查询：“台积电的产能和技术”。
    - **左侧结果区:**
      * **标题:** Graph Search Results (结构化知识)
      * **内容:** 显示从Neo4j中找到的、与“台积电”直接相关的事件，如“竹南先进封测六厂(AP6)已正式启用...”。
    - **右侧结果区:**
      * **标题:** Vector Search Results (语义化知识)
      * **内容:** 显示从ChromaDB中找到的、语义上与“产能”、“技术”相关的事件或原始文本片段。
    - **中间:** 一个加号图标，表示两种结果的融合。
* **演讲者备注:**
  + 解释两种搜索方式的互补性：图搜索保证了结果的准确性和逻辑性，向量搜索保证了结果的广度和相关性。
  + 强调这种混合搜索是后续生成高质量“风险分析思维链”的基础。

**Slide 12: 核心应用：风险分析思维链 [新增/强化页]**

* **标题:** 核心应用：生成可解释的风险分析思维链
* **核心信息:** 展示系统的最终产出——基于事理超图的、可解释的、多步骤的风险传导路径。
* **视觉设计:**
  + **思维链图 (Chain of Thought Visualization):**
    1. **起点 (外部事件):** 一个事件框，“日本宣布对尖端半导体设备实施出口管制”。
    2. **第一跳 (图谱内事实推断):** 箭头指向下一个框，“**[事实依据]** RelationshipAnalysisAgent 基于图谱中已有的‘A公司依赖日本设备’的事实，推断：**[直接影响]** A公司的设备采购将受阻。”
    3. **第二跳 (领域知识推理):** 箭头指向下一个框，“**[领域模型]** LLM 结合‘设备是产能扩张前提’的领域知识，进一步推理：**[潜在风险]** A公司先进制程的产能扩张计划可能延后。”
    4. **第三跳 (下游连锁风险):** 箭头指向最终框，“**[下游关联]** 基于图谱中‘B公司是A公司芯片的主要客户’的关系，推断：**[连锁风险]** B公司的AI服务器产品线可能面临供应链风险。”
  + **标注:** 在每个箭头上标注推理的类型（事实推断、领域推理、关联推断），增强可解释性。
* **演讲者备注:**
  + 这是整个报告的高潮，展示了系统如何模拟专家的思考过程。
  + 强调“思维链”的每一步都是有据可查、可解释的，这与黑箱模型有本质区别。

**Slide 13: 结论与未来展望**

* **标题:** 结论与未来展望
* **页面布局:** 左右分栏。
* **左栏：研究结论**
  + **内容:**
    1. 提出并实现了一套**新颖的三阶段框架**，有效融合了超图与大语言模型。
    2. 设计的 **Cortex 引擎** 和 **知识闭环机制**，显著提升了知识图谱构建的自动化程度与知识质量。
    3. 验证了基于高保真度事理超图生成**可解释风险分析思维链**的可行性与有效性。
* **右栏：未来展望**
  + **内容:**
    1. **算法层面:** 探索基于扩散模型的**图纯化与补全**方法，进一步提升图谱质量。
    2. **应用层面:** 开发面向信审、投研等场景的**可视化交互分析系统**。
    3. **领域拓展:** 将该框架应用于**生物医药、法律文书**等其他知识密集型领域。
* **演讲者备注:**
  + 有力地总结本研究的核心贡献，并为未来的研究工作指出清晰的方向。

**Slide 14: Q&A [新增/强化页]**

* **标题:** 感谢聆听，敬请指正
* **视觉设计:** 页面简洁，居中放置“Q&A”，下方是您的联系方式（如邮箱）和项目Github链接（如果有）。页眉页脚保留校徽。