

这份讲座由 IBM Research 的 **Panos** 和 IBM CIO 组织的 **Abby** 共同主讲，主题是**“使用 Docling 解锁文档智能”(Unlock Document Intelligence with Docling)**。

这是整个系列的最后一部分，重点解决了一个 AI 领域最基础但也最棘手的问题：数据摄入 (**Data Ingestion**)。无论你的 Agent 或 RAG 系统有多先进，如果它无法正确理解输入的 PDF 或文档格式，效果就会大打折扣。

以下是讲座的核心内容总结：

1. 核心问题：PDF 的陷阱

- 非结构化数据难题：企业中大量的知识被“锁”在 PDF、PPT 和 Word 文档中。
- 传统解析的失败：传统的 PDF 解析器（如 PyPDF）通常只是提取纯文本，忽略了布局 (**Layout**) 和结构 (**Structure**)。
 - 案例：Panos 展示了一个经典错误。科学论文中的标题 "Vegetative Electron Microscopy" 被分在两列。简单解析器会跨列读取，将其读成无意义的 "Vegetative Electron" 和 "Microscopy"，完全破坏了语义。
 - 后果：这导致了 RAG 系统中的“垃圾进，垃圾出”(Garbage In, Garbage Out)。

2. 解决方案：Docling

- 简介：Docling 是 IBM Research 推出的一个开源 **Python** 库，专为 AI 时代设计，用于将复杂的文档转换为结构化的、AI 可读的格式（如 Markdown, JSON）。
- 核心能力：
 - 高精度布局分析：它不仅仅看文本，还使用视觉模型 (Vision Models) 来“看”文档，能够识别标题、段落、页眉页脚。
 - 表格结构识别：能够处理复杂的跨行、跨列表格，并将其转换为结构化数据，这对金融和科研文档至关重要。
 - 阅读顺序恢复：智能识别多栏排版，确保按人类阅读顺序提取文本。
 - 多模态支持：内置 OCR，可以处理扫描件和图片中的文本。

3. 技术亮点与新发布

- **Granite Docling 模型**：宣布了基于 Granite 的新视觉语言模型 (VLM)，专门优化用于文档提取，体积小但性能强大，可以在普通 GPU 甚至 CPU 上运行。
- **Docling Serve**：推出了服务器端运行模式，支持 API 调用。

- **Docling MCP:** Docling 现在也可以作为 **MCP (Model Context Protocol)** 服务器运行。这意味着你可以像之前的演示一样，将 Docling 直接连接到 Claude 或其他 Agent 中，让 AI 直接具备“阅读复杂文档”的能力。

4. 实战案例：IBM CIO 内部应用 (AskIBM)

- **Abby 的分享：**作为 IBM 内部 IT 系统的负责人，她分享了 Docling 如何在 "AskIBM" (IBM 的内部 AI 助手) 中落地。
- **挑战：**内部知识库包含海量的历史文档、政策文件和技术手册，格式极其混乱。
- **应用：**使用 Docling 建立了一个标准化的数据摄入管道。
 - **效果：**通过保留文档的层级结构 (Hierarchy)，RAG 检索的准确率大幅提升。员工提问时，AI 能精确定位到具体的章节或表格单元格，而不是给出模糊的答案。

5. 总结

Docling 填补了**“人类可读文档”与“机器可理解数据”**之间的鸿沟。它不仅仅是一个转换工具，更是构建高质量 RAG 和 Agent 应用的基础设施。只有当 AI 能够像人类一样理解文档的排版和结构时，自动化的价值才能真正释放。