

How Box Evolved from Simple AI to Agentic Systems for Enterprise | LangChain Interrupt



LangChain
134K subscribers

Subscribe

80



Share

Clip

Save



4,640 views Jun 10, 2025

Box's CTO, Ben Kus, shares how their AI document extraction system hit a wall when customers brought real-world complexity—300-page documents, 500+ data fields, and OCR challenges that broke their simple AI approach. After initially deploying to 10 million pages successfully, Box completely re-architected their system using a multi-agent agentic approach that separated complex problems into specialized sub-agents. This transformation not only solved their scaling issues but made their system dramatically easier to evolve and update for enterprise needs.

Watch all of our recorded sessions from Interrupt here: <https://interrupt.langchain.com/video...>

UNLOCKING AGENT CREATION: AGENTIC ARCHITECTURE LESSONS LEARNED AT BOX



67%

Fortune 500

115K

Total customers



Box has fully integrated generative AI starting in 2023

Q&A across documents



How can we speed up our product roadmap?

There are several delays in the product roadmap that I will highlight:

1. Your roadmap schedule includes a sequential QA and legal review process. Consider running these two processes concurrently to save review time.
2. Your product strategy includes customer focus groups for multiple features. Consider running one larger focus group.

Data extraction

172 Contracts

17 Above \$5,000

90 Due in 90 Days

12 Risky clauses

23 Out for signature

53 Contain PII

AI-powered workflows



Loan Approved

DUE IN 7 DAYS

APPROVED

Signature required

New client onboarded

JOHN SMITH | JSMITH@ACME.COM



Insurance claim processed

PENDING

Documents required



Compliance audit submitted

CONFIDENTIAL

Box has fully integrated generative AI starting in 2023

Q&A across documents



How can we speed up our product roadmap?

There are several delays in the product roadmap that I will highlight:

1. Your roadmap schedule includes a sequential QA and legal review process. Consider running these two processes concurrently to save review time.
2. Your product strategy includes customer focus groups for multiple features. Consider running one larger focus group.

Data extraction

172 Contracts

17 Above \$5,000

90 Due in 90 Days

12 Risky clauses

53 Contain PII

AI-powered workflows



Loan Approved

DUE IN 7 DAYS

APPROVED

Signature required

New client onboarded

JOHN SMITH | JSMITH@ACME.COM



Insurance claim processed

PENDING

Documents required



Compliance audit submitted

CONFIDENTIAL

Metadata extraction enables AI to understand and process enterprise content effectively



Powering AI with context-rich content



Automating workflows with intelligent content understanding



Transforming content discovery and knowledge Access

Non-Disclosure Agreement (NDA)

(V01-01-2024)

This Non-Disclosure Agreement or ("Agreement") is by and between:

Party Disclosing Information: ACME Inc. with a mailing address of 123 Company Lane, San Francisco, CA ("Disclosing Party").

Party Receiving Information: Galactech Inc. ("Receiving Party").

The expiration date of this agreement is November 12 2025.

For the purpose of preventing the unauthorized disclosure of Confidential Information as defined below, the parties agree to enter into a confidential relationship concerning the disclosure of certain proprietary and confidential information ("Confidential Information").

1. **Definition of Confidential Information.** For purposes of this Agreement, "Confidential Information" shall include all information or material that has or could have commercial value or other utility in the business in which Disclosing Party is engaged. If Confidential Information is in written form, the Disclosing Party shall label or stamp the materials with the word "Confidential" or some similar warning. If Confidential Information is transmitted orally, the Disclosing Party shall promptly provide writing indicating that such oral communication constituted Confidential Information.

2. **Confidentiality Obligations.** Each Party agrees that all Confidential Information provided by the other Party is proprietary and confidential, including but not limited to trade secrets, business methodologies, financial data, client information, and any other sensitive material disclosed during the relationship. Confidential Information shall only be used for the intended purpose of this Agreement and shall not be disclosed to any third parties without the express written permission of the Disclosing Party, except to employees, agents, or subcontractors who are under comparable obligations of confidentiality with industry standards and as required by the Personal Information Protection and Electronic

Metadata

Contract

Contract Type: NDA

Version: V01-01-2024

Participant: Galactech Inc.

Number of Parties: 2

Expiration Date: 11-12-2025

Over 90% of enterprise data is unstructured, with critical metadata buried deep within

Before GenAI, extracting data was too hard for most enterprises



ML-based techniques too expensive to develop



Didn't work well for free form text data



Not responsive to data format changes

Building generative AI data extraction First version



Performed well

Flexible, based on reasoning capabilities of LLMs

Built around a specific AI model

Required ML-based pre-processing for OCR

Building generative AI data extraction Challenges



Started getting longer / more complex docs

OCR would sometimes lose valuable info

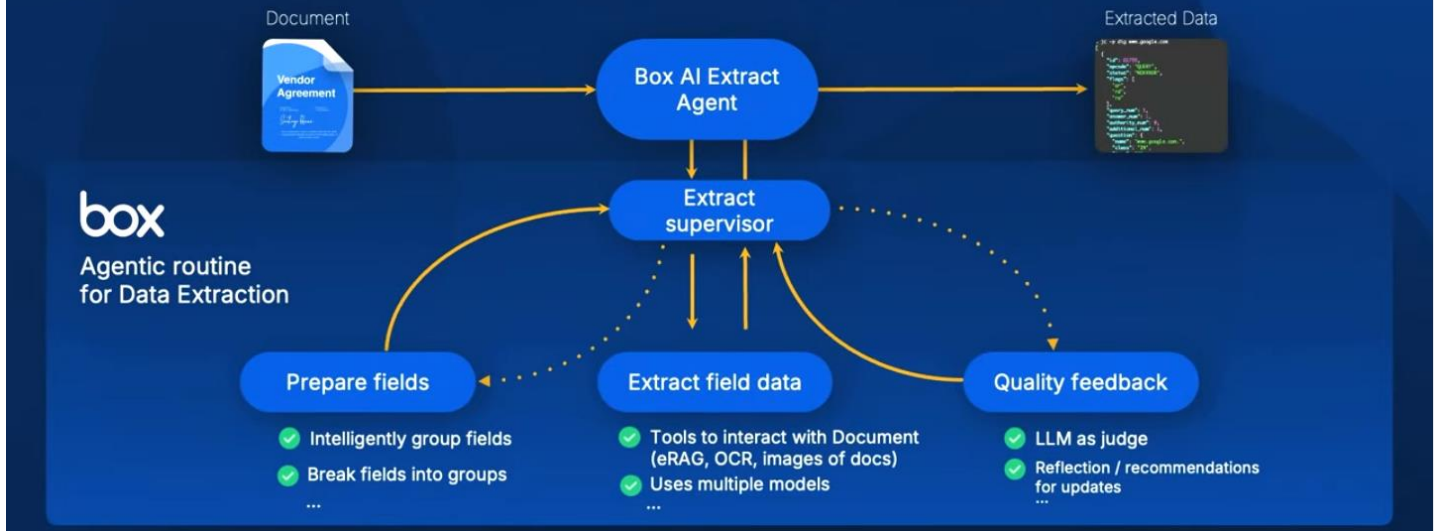
Too many fields would overwhelm the LLM

LLM-as-a-judge could determine that sometimes it was wrong on its first pass

Our simple, elegant LLM-powered architecture became convoluted, fragile, and hard to update



Rearchitecting for agentic approach



Lesson learned

Build agentic architecture early!

Cleaner platform abstraction

Separation of high-scale infra work vs. improving specific agent actions

Ease to evolve quickly

Easily adapt as new techniques and AI models become available

Improved engineering AI culture

Helps engineers think agentially (state diagrams, Tools, A2A, ...) as part of overall ecosystem