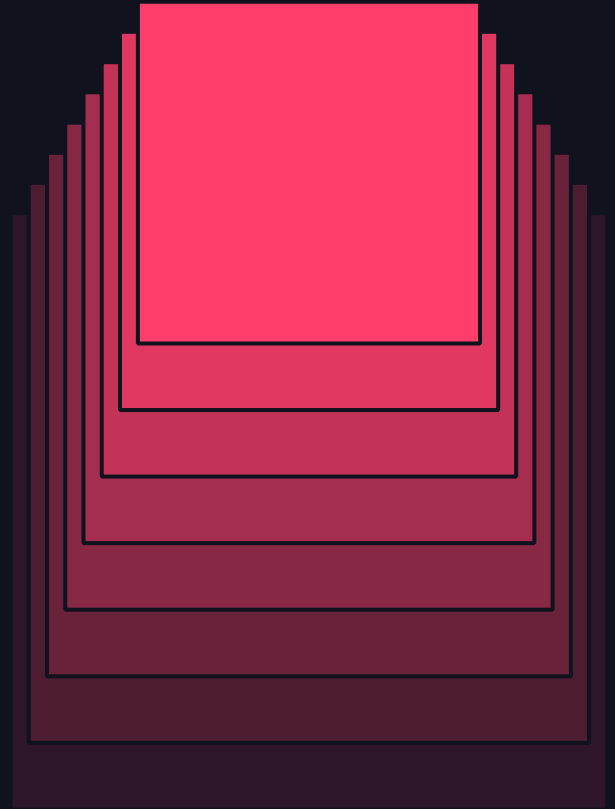


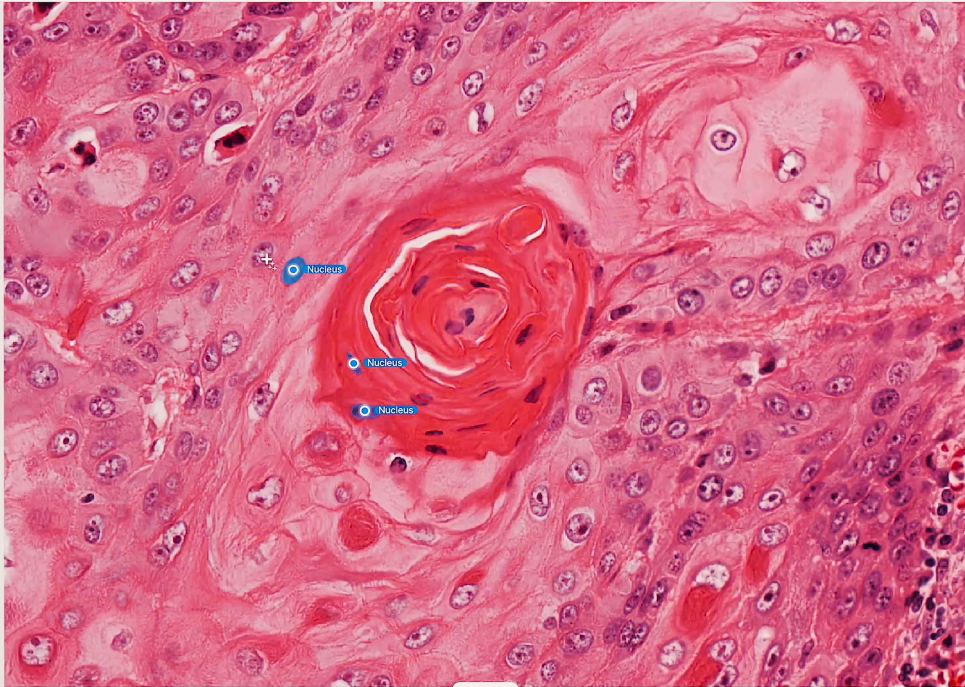
# Lakehouse to LLM: Automating Knowledge work at scale

---

Alberto Rizzoli – CEO, V7  
June 2024



Segment Anything... Nucleus 1 / 87 04:04 Discard Saving 1



Annotations Review

- Nucleus
- Nucleus
- Nucleus

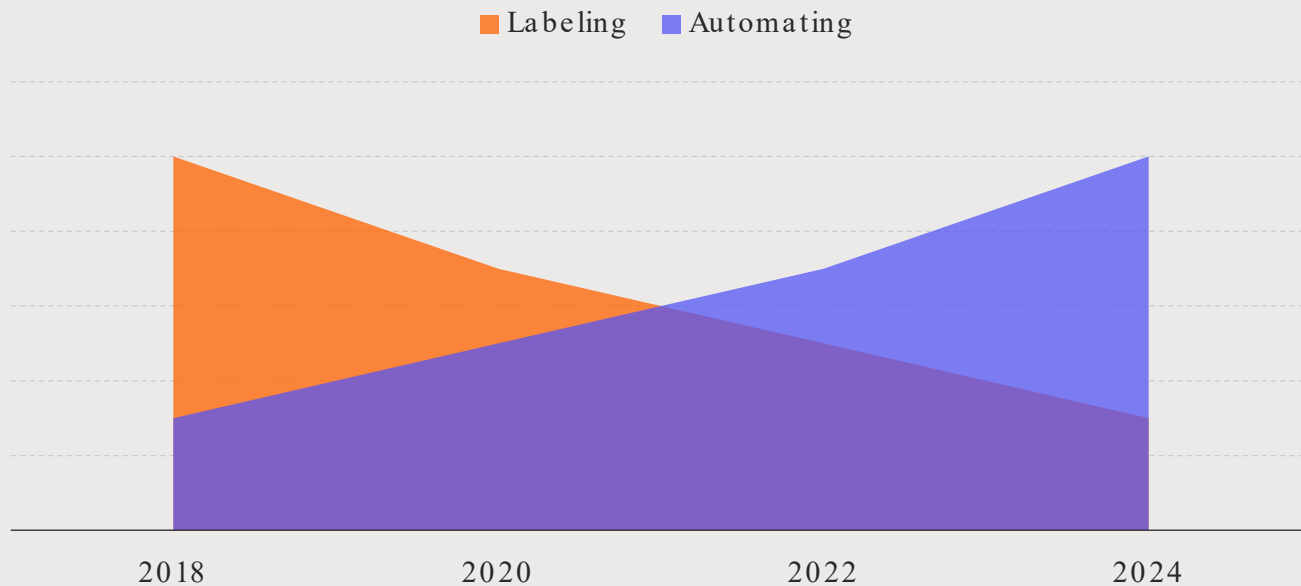
Tags 14

No tags available

Type to search or add a tag

All 87

# What AI leadership should spend time on



What do you call it  
when employees  
produce ground  
truth?

How do we capture knowledge work and  
bring it into LLMs?

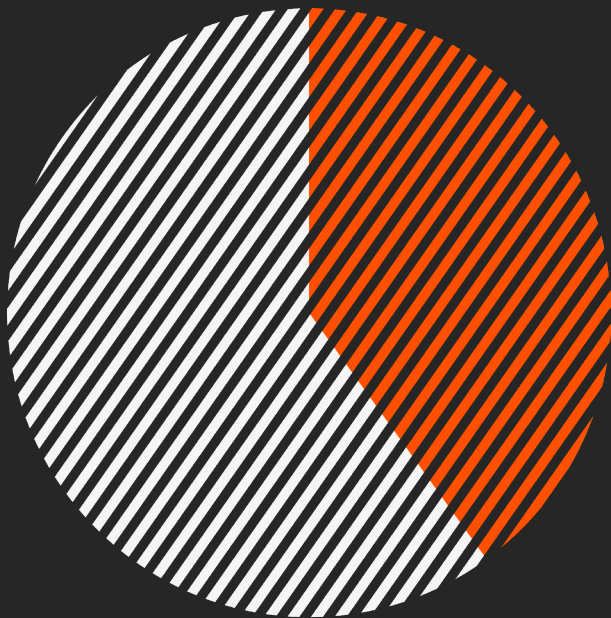






Chapter 1

# Designing for knowledge work automation



40%

of knowledge work can be done by well-trained frontier models just as well.

---

So what's missing? Why aren't we doing it yet?



# Designing to adapt to any form of knowledge work

## Format support

Must adopt text, images, audio calls, documents, natural language instructions.

## Discrete decisions

Must convert the output of LLMs into discrete options, easier to set up than function calls.

## Reliability at Scale

After completing 1,000 reviewed tasks it must succeed and fail reliably.

# What do companies care about?

## Compounding improvements

Investing in AI should improve the business, not its model provider.

## Freedom to build

Not be locked into restrictive domain-specific SaaS or long model implementations.

## Choice of model

If the company already invested into one particular provider.



x ^ v

🔒

**Invoice**

**Invoice no: 40378170**  
Date of issue: 10/15/2012

**Seller:** Patel, Thompson and Montgomery  
356 Kyle Vista  
New James, MA 46228  
Tax Id: 958-74-3511  
IBAN: GB77WRBQ31965128414006

**Client:** Jackson, Odonnell and Jackson  
267 John Track Suite 841  
Jenniferville, PA 98601  
Tax Id: 998-87-7723

**ITEMS**

No.	Description	Qty	UM	Net price	Net worth	VAT (%)	Gross worth
1	Leef's Wine Companion Bottle Corkscrew Opener Gift Box Set with Foil Coater	1.00	each	7.50	7.50	10%	8.25

**SUMMARY**

	VAT (%)	Net worth	VAT	Gross worth
	10%	7.50	0.75	8.25
<b>Total</b>		<b>\$ 7.50</b>	<b>\$ 0.75</b>	<b>\$ 8.25</b>

**OCR**

Invoice no: 40378170  
Date of issue:  
Seller:  
Patel, Thompson and Montgomery  
356 Kyle Vista  
New James, MA 46228  
Tax Id: 958-74-3511  
IBAN: GB77WRBQ31965128414006  
ITEMS  
10/15/2012  
Client:  
Jackson, Odonnell and Jackson  
267 John Track Suite 841  
Jenniferville, PA 98601  
Tax Id: 998-87-7723  
No.  
Description  
Qty  
UM  
Net price  
Net worth  
VAT 1%

**State**

Search

- Pennsylvania
- Oregon
- Colorado
- New York
- Illinois
- Connecticut
- Tennessee
- Louisiana

**Amount**

\$8.25

**Address**

267 John Track Suite 841, Jenniferville, PA 98601

**Mistakes**

Search

- No mistake
- Mistake

**IBAN**

GB77 WRBQ 3196 5128 4140 06

**Date**

15/10/2012

# Discrete Decisions

Export data

Ask Go

description of product	Size	Brand
Red Bull Sugarfree...	Energy Drinks	Red Bull
Head & Shoulders ...	Hair Care	Procter & Gamble
Pompeian Balsam ...	Oils and Vinegars	Pompeian
Ziploc Heavy Duty ...	Kitchenware	Ziploc
Kellogg's Original ...	Snacks	Kellogg's
Kellogg's Original ...	Snacks	Kellogg's
Lindt Excellence ...	Chocolate	Lindt
Lindt Excellence ...	Chocolate	Lindt
Kraft Original Cheddar ...	Condiments	Kraft
Kraft Original Cheddar ...	Condiments	Kraft
Nabisco Original Cheddar ...	Snacks	Nabisco
Nabisco Original Cheddar ...	Snacks	Nabisco
Coca-Cola Life ...	Beverages	Coca-Cola
Coca-Cola Life ...	Beverages	Coca-Cola

Spend category    GPT-4 or Gemini    +

Spend category

Type    Multi Select >

Tool    GPT 4 >

Inputs    Select inputs >

Assume the spend category of this particular product, based on its brand name, average price point, and product positioning.

Select one of the three:

- Essentials: For low-end, economical product with no frills

Q Search or add options

- Essentials
- Mid-Range
- High-end

Recompute

# Choice of Model

Ask Go

Amount	Address	Mistakes	IBAN
98601		No mistake	GB77 WRBQ 3196 5128 4140 06
		No mistake	GB82 DJFU 8227 3146 2104 25
		No mistake	GB32 AMZY 3134 0900 4611 45
0 37508		No mistake	GB36 GFUF 9334 2035 3549 16
		No mistake	GB69 HUXE 8246 0947 3164 50
		No mistake	GB44 IJGO 6410 3154 1906 92
98601		No mistake	GB77 WRBQ 3196 5128 4140 06
, LA 32852		No mistake	GB77 WRBQ 3196 5128 4140 06

Amount

Type ☰ Text >

Tool 🌐 GPT 3.5 Turbo >

Inputs 📄 Invoice >

What's the total amount of the invoice in dollars? Just return the value in the following format: \$212.09

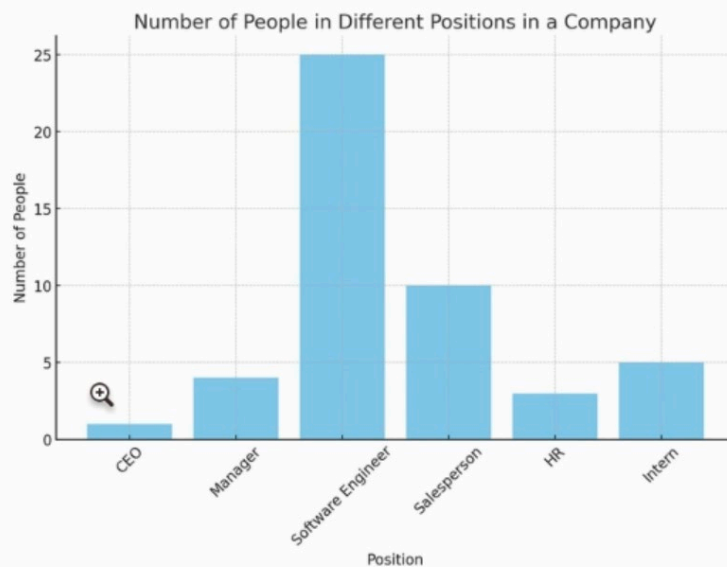


## Historical and Projected Adj. EBITDA

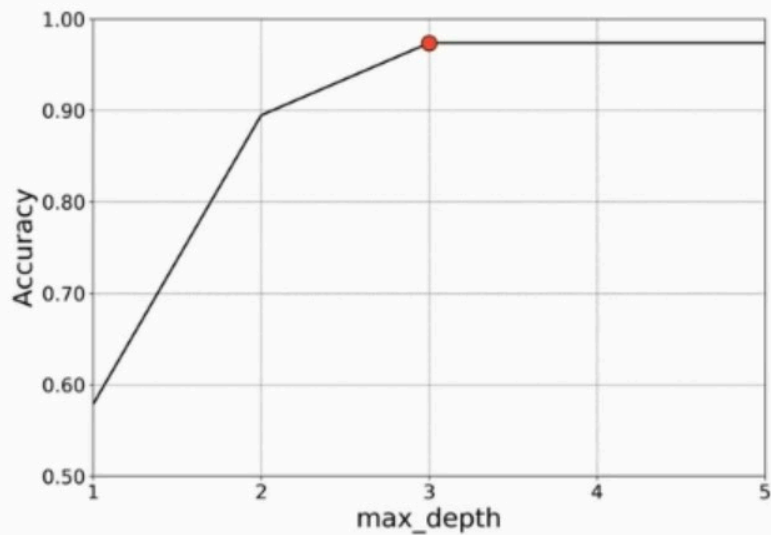
(\$ in millions)



Year	Adj. EBITDA (Gemini)	Adj. EBITDA Margin (Gemini)	Adj. EBITDA (Omni)	Adj. EBITDA Margin (Omni)	Adj. EBITDA (Opus)	Adj. EBITDA Margin (Opus)	Adj. EBITDA (Turbo)	Adj. EBITDA Margin (Turbo)
2020	\$7.1	12.9%	\$7.1	12.9%	\$7.1	12.9%	\$7.1	12.9%
2021	\$7.1	12.6%	\$7.1	12.6%	\$7.1	12.6%	\$7.1	12.6%
2022E	\$11.9	13.4%	\$11.9	13.4%	\$11.9	13.4%	\$11.9	13.4%
2023P	\$11.9	13.0%	\$11.9	13.0%	\$11.9	13.0%	\$11.9	13.0%
2024P	\$13.0	13.2%	\$13.0	13.2%	\$13.0	13.2%	\$13.0	13.2%
2025P	\$13.4	13.3%	\$13.4	13.3%	\$13.4	13.3%	\$13.4	13.3%



Position	Gemini 1.5	Omni	Opus	Turbo	Ground Truth
CEO	1	1	1	2	1
Manager	4	5	20	5	4
Software Engineer	25	25	15	23	25
Salesperson	10	15	10	8	10
HR	3	4	5	4	3
Intern	5	6	3	6	5
Accuracy	100%	83.3%	50%	66.7%	100%



I choose max\_depth =3 as it seems to be an accurate model and not the most complicated.

max_depth	<b>Gemini</b> Accuracy	<b>Omni</b> Accuracy	<b>Opus</b> Accuracy	<b>Turbo</b> Accuracy
1	0.60	0.60	0.90	0.50
2	0.90	0.85	0.70	0.70
3	0.97	0.95	0.80	0.90
4	0.97	0.95	0.70	0.95
5	0.97	0.95	0.70	0.97

# Reliability at Scale



# Prompting

Start with this

..keep trying...

# Fine Tuning

If you really have to



	☰ Deal value	🗒 AE	🕒 Follow up	📁 Needed Paperwork	+	...
...	\$40000	Katherine	Client needs to follow up to	📁 1 collection item	•	
er Deal	\$20000	Katherine	Client needs to follow up to	📁 1 collection item	•	
Content ...	\$200,000	Harry	Client needs to follow up to	📁 1 collection item	•	
Vision ...	300000	• Harry	Client needs to follow up to	📁 1 collection item	•	
Sensitive 🗒	200000	• Harry	AE needs to follow up	📁 3 collection items	•	
Go image...	30000	• Katherine	Client needs to follow up to	📁 3 collection items	•	
nz Supply ...	30000	• Katherine	Client needs to follow up to	📁 1 collection item	•	
thineers - ...	100000	• Katherine	Client needs to follow up to	📁 3 collection items	•	
ns - Go De...	100000	• Will	We are awaiting paperwork	📁 3 collection items	•	
24Q2 - GO...	30000	• Harry	AE needs to follow up	📁 3 collection items	•	
4Q2 - ...	60000	• Harry	Client needs to follow up to	📁 1 collection item	•	
s/03/2024 ...	80000	• Jason	AE needs to follow up	📁 3 collection items	•	
orm & ...	30000	• Harry	Client needs to follow up to	📁 3 collection items	•	
no - ...	30000	• Katherine	AE needs to follow up	📁 1 collection item	•	
t & Skelet...	40000	• Katherine	Client needs to follow up to	📁 2 collection items	•	



# Context

System Prompt - Avoid preamble, format this way

Task Prompt - Does the liability clause raise any issues?

Inputs - Target document, guidelines document

Ground truth samples - "10m liability cap is ok"

Task Context - There is/isn't a human supervising

# Dull, Detailed, Growing

Can we model repetitive knowledge work as unstructured data tasks, and solve them with LLMs?



## Finance

- CIMs
- 10k forms
- Quarterly Reports
- Tax Forms, Audits



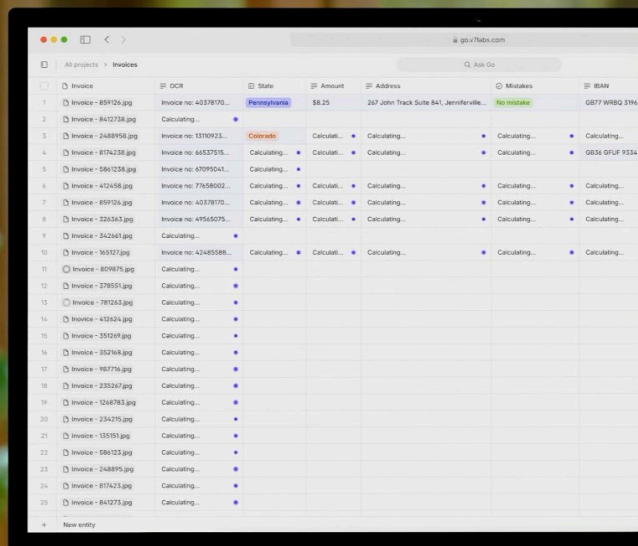
## Logistics

- Bills of Lading
- Email Comms
- Packing lists
- Transit information



## RLHF and Labeling

- Side by Side workflows
- Human Consensus
- Automatic AI Reviews



	Invoice	OCR	State	Amount	Address	Mutates	IBAN
1	Invoice - 859126.jpg	Invoice no: 40378170...	Pennsylvania	\$8.25	267 John Track Suite B41, Jenneville...	No mutates	GB77 WIR0 3194
2	Invoice - 8412758.jpg	Calculating...					
3	Invoice - 2488958.jpg	Invoice no: 13110923...	Colorado	Calculat...	Calculating...	Calculating...	Calculating...
4	Invoice - 8174238.jpg	Invoice no: 66537515...		Calculating...	Calculat...	Calculating...	GB36 GFLF 9314
5	Invoice - 5861238.jpg	Invoice no: 670995041...		Calculating...			
6	Invoice - 472458.jpg	Invoice no: 77658002...		Calculating...	Calculat...	Calculating...	Calculating...
7	Invoice - 859126.jpg	Invoice no: 40378170...		Calculating...	Calculat...	Calculating...	Calculating...
8	Invoice - 326165.jpg	Invoice no: 49565075...		Calculating...	Calculat...	Calculating...	Calculating...
9	Invoice - 342661.jpg	Calculating...					
10	Invoice - 56317.jpg	Invoice no: 42485588...		Calculating...	Calculat...	Calculating...	Calculating...
11	Invoice - 809875.jpg	Calculating...					
12	Invoice - 378551.jpg	Calculating...					
13	Invoice - 787265.jpg	Calculating...					
14	Invoice - 432424.jpg	Calculating...					
15	Invoice - 353189.jpg	Calculating...					
16	Invoice - 352168.jpg	Calculating...					
17	Invoice - 982779.jpg	Calculating...					
18	Invoice - 235267.jpg	Calculating...					
19	Invoice - 5268763.jpg	Calculating...					
20	Invoice - 234715.jpg	Calculating...					
21	Invoice - 155193.jpg	Calculating...					
22	Invoice - 586125.jpg	Calculating...					
23	Invoice - 248895.jpg	Calculating...					
24	Invoice - 817423.jpg	Calculating...					
25	Invoice - 841273.jpg	Calculating...					
+	New entry						

## Chapter 2

# Building our first Document automation workflow

# Logistics:

A world of  
necessary paperwork  
and unnecessary  
manual work

**Anonymized Demo Document**  
**MULTIMODAL TRANSPORT OR PORT TO PORT SHIPMENT**

Bill of Lading

\*Waybill     \*Bill of Lading

SHIPPER/EXPORTER INCREDIBLE FURNITURE CO., LTD O/B: LIMITED LAND LOT NO.111, MAP NO.1, TAO QUARTER, WAN COMMUNE, BINH DUONG PROVINCE, VIETNAM	DOCUMENT NO. SGN1040879	BILL NUMBER 220221320T
EXPORT REFERENCES CUST PO0207149 CONS		
CONSIGNEE AMERICAN DEMO FREIGHT INC. 1200 WASHINGTON AVE - STE: 200 DALLAS TX 75006	FORWARDING AGENT-REFERENCES INTERPOST VIETNAM COMPANY 18A RAO STREET, WARD 1 TAN BINH DISTRICT HO CHI MINH CITY VN	
POINT AND COUNTRY OR TERRITORY OF ORIGIN HO CHI MINH CITY VN		
NOTIFY PARTY UNITED FURNITURE INDUSTRIES, INC 5380 HWY 145 SOUTH TUPELO, MS 38801 ATTN: BELLENA E. CRIDDLE 662-432-0431	DELIVERY AGENT INTERPOST OCEAN, INC. FMC #0001 2700 YELLOW RED ROAD BLDG M, SUITE 300 DALLAS TX 77032 US	
PLACE OF RECEIPT HO CHI MINH CITY	FCL	
EXPORTING CARRIER (Vessel Name or Designation of Other Means of Transportation) CMA CGM A. LINCOLN 0TUGRE1MA	SEA/AIR PORT OF LOADING VUNG TAU	FINAL DESTINATION OCEAN PORT To RAMP
SEA/AIR PORT OF DISCHARGE LOS ANGELES, CA	PLACE OF DELIVERY DALLAS, TX	PRE-CARRIAGE BY PLACE OF RECEIPT BY PRE-CARRIER

**PARTICULARS FURNISHED BY SHIPPER**

MARKS AND NUMBERS	NO. OF PKGS.	DESCRIPTION OF PACKAGES AND GOODS (see Paragraphs 15-16 on reverse side)	GROSS WEIGHT	MEASUREMENT
CTRNO: TCNU3349111 SEALCODE1: H3326111	1	40' HIGH CUBE CONTAINER 80 CARTON SAID TO CONTAIN PO: 110159	4,967.200 KG 10,950.800 LB	64.970 M3 2,294.091 F3

**Anonymized Demo Document**  
**MULTIMODAL TRANSPORT OR PORT TO PORT SHIPMENT**

Bill of Lading

\*Waybill  \*Bill of Lading

SHIPPER/EXPORTER INCREDIBLE FURNITURE CO., LTD O/B: LIMITED LAND LOT NO.111, MAP NO.1, TAO QUARTER, WAN COMMUNE, BINH DUONG PROVINCE, VIETNAM		DOCUMENT NO. SGN1040879	BILL NUMBER 220221320T
		EXPORT REFERENCES CUST PO0207149 CONS	
CONSIGNEE AMERICAN DEMO FREIGHT INC. 1200 WASHINGTON AVE - STE: 200 DALLAS TX 75006		FORWARDING AGENT-REFERENCES INTERPOST VIETNAM COMPANY 18A RAO STREET, WARD 1 TAN BINH DISTRICT HO CHI MINH CITY VN	
		POINT AND COUNTRY OR TERRITORY OF ORIGIN HO CHI MINH CITY VN	
NOTIFY PARTY UNITED FURNITURE INDUSTRIES, INC 5380 HWY 145 SOUTH TUPELO, MS 38801 ATTN: BELLENA E. CRIDDLE 662-432-0431		DELIVERY AGENT INTERPOST OCEAN, INC. FMC #0001 2700 YELLOW RED ROAD BLDG M, SUITE 300 DALLAS TX 77032 US	
PLACE OF RECEIPT HO CHI MINH CITY		FCL	
EXPORTING CARRIER (Vessel Name or Designation of Ship, Train or Transportation) CMA CGM A. LINCOLN 0TUGRE1MA	SEA/AIR PORT OF LOADING VUNG TAU	FINAL DESTINATION OCEAN PORT To RAMP	
SEA/AIR PORT OF DISCHARGE LOS ANGELES, CA	PLACE OF DELIVERY DALLAS, TX	PRE-CARRIAGE BY	PLACE OF RECEIPT BY PRE-CARRIER

**PARTICULARS FURNISHED BY SHIPPER**

MARKS AND NUMBERS	NO. OF PKGS.	DESCRIPTION OF PACKAGES AND GOODS (see Paragraphs 15-16 on reverse side)	GROSS WEIGHT	MEASUREMENT
CTRNO: TCNU3349111 SEALCODE1: H3326111	1	40' HIGH CUBE CONTAINER 80 CARTON SAID TO CONTAIN PO: 110159 CUS.PO: PO1101149 HS CODE: 940111 BEDROOM FURNITURE TSCA TITLE VI & CARB PHASE 2 COMPLIANT THIS SHIPMENT CONTAINS NO SOLID WOOD PACKING MATERIAL	4,967.200 KG 10,950.800 LB	64.970 M3 2,294.091 F3

Large Logistics Provider

300,000/day

500+ variants

Handwriting

12 Languages

\$50m/y to process

Recent Projects

+ New Project

Invite Users

All Projects

- Lemon Terabyte Green
- Uninterested Vegetable Easy
- Hallowed Apple Important
- Broken Cat Bent
- Yellow Sandwich Shy
- Reviews
- Invoices
- Red Insect Red
- Digital Pathology
- LinkedIn Pages
- Competitor Research
- Document Verification
- PDF Splitter

- LT** **Lemon Terabyte Green**  
0 properties
- UV** **Uninterested Vegetable Easy**  
6 properties
- HA** **Hallowed Apple Important**  
0 properties
- BC** **Broken Cat Bent**  
4 properties
- YS** **Yellow Sandwich Shy**  
8 properties
- Re** **Reviews**  
6 properties



What's Next?



Library

Build

Review

Automate

Export

File - Drop any file

1 c9aff653-1a8d-4d29...

2 a7310f8d-0456-4d2...

3 f2e31436-6a08-42af...

4 74d35584-99e4-48...

5 deck00009.pdf

6 deck00008.pdf

7 deck00000.pdf

+ New entity



### Hi, this is Ask Go.

AskGo is an assistant that can understand your project and take actions. Building your project with AskGo is as simple as asking a question. How can AskGo help you today?

Extract line items from receipts

Set up an RLHF workflow

Classify emails and collect attachments

Ask Go...



[alberto@v7labs.com](mailto:alberto@v7labs.com)  
We're hiring

[V7labs.com/go](https://v7labs.com/go)