INDEPENDENT RESEARCH IN DATA SCIENCE

# Entity Extraction Strategies for Historical Mortgage Records

By Uday Sapra

Under Kate Thomas and Jacob Faber (NYU Redlining Lab)

Dec 12, 2024

# Agenda

- What was the research about?
- What was my role?
- What has been the output?
- What has been learnt?
- What is the approach going forward?

01 The Team & Research

02 Literature Review

03 Incompatibility with our Data

04 Renewed Approach

05 Outputs and Validation

06 Conclusion and Next Steps

NYU REDLINING LAB

The Research

# Data-Oriented Approach to Delineating the Long Term Impacts of Government Redlining

- **Redlining in Real Estate:** Denying access to credit (mortgage and refinancing) based on race
- **Impact:** Segregation, inequalities in neighbourhood development, wealth creation, opportunities, & asset values

**My Role?**

**Create the Process to build the Dataset that enables this research - from Raw Document Scans**

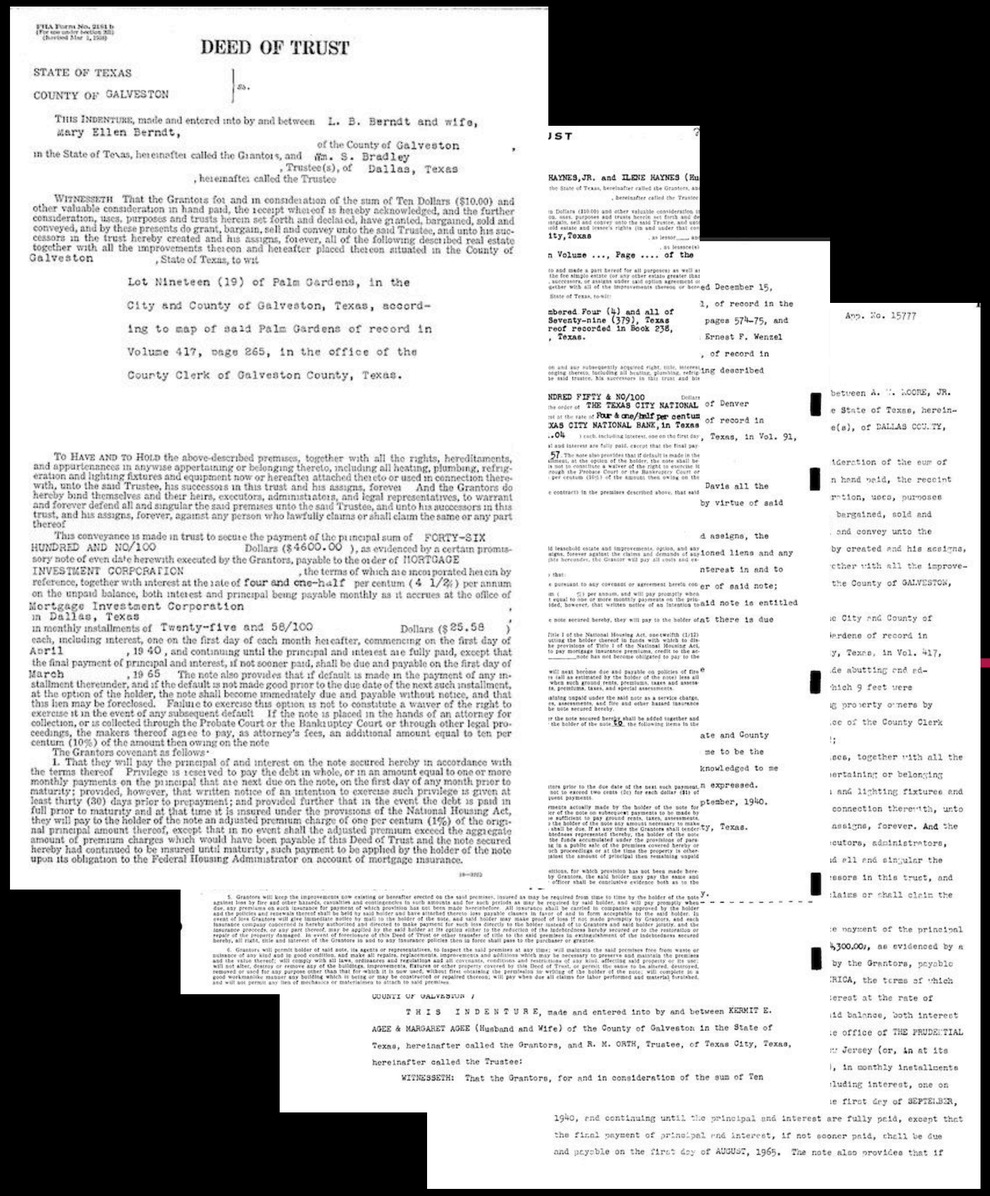
## OUR DATA

**280,000**
Mortgage Document Scans (.TIF)

**20**
Diverse Counties

**1930-1975**
Four Decades Spanned

# Document Corpus



DEED OF TRUST

STATE OF TEXAS
COUNTY OF GALVESTON

THIS INDENTURE, made and entered into by and between L. B. Berndt and wife, Mary Ellen Berndt, of the County of Galveston in the State of Texas, hereinafter called the Grantors, and Wm. S. Bradley, Trustee(s), of Dallas, Texas, hereinafter called the Trustee

WITNESSETH That the Grantors for and in consideration of the sum of Ten Dollars ($10.00) and other valuable consideration in hand paid, the receipt whereof is hereby acknowledged, and the further consideration, uses, purposes and trusts herein set forth and declared, have granted, bargained, sold and conveyed, and by these presents do grant, bargain, sell and convey unto the said Trustee, and unto his successors in the trust hereby created and his assigns, forever, all of the following described real estate together with all the improvements thereon and hereafter placed thereon situated in the County of Galveston, State of Texas, to wit

Lot Nineteen (19) of Palm Gardens, in the City and County of Galveston, Texas, according to map of said Palm Gardens of record in Volume 417, page 265, in the office of the County Clerk of Galveston County, Texas.

TO HAVE AND TO HOLD the above-described premises, together with all the rights, hereditaments, and appurtenances in anywise appertaining or belonging thereto, including all heating, plumbing, refrigeration and lighting fixtures and equipment now or hereafter attached thereto or used in connection therewith, unto the said Trustee, his successors in this trust and his assigns, forever. And the Grantors do hereby bind themselves and their heirs, executors, administrators, and legal representatives, to warrant and forever defend all and singular the said premises unto the said Trustee, and unto his successors in this trust, and his assigns, forever, against any person who lawfully claims or shall claim the same or any part thereof

This conveyance is made in trust to secure the payment of the principal sum of FORTY-SIX HUNDRED AND NO/100 Dollars ($4600.00), as evidenced by a certain promissory note of even date herewith executed by the Grantors, payable to the order of MORTGAGE INVESTMENT CORPORATION, the terms of which are incorporated herein by reference, together with interest at the rate of four and one-half per centum (4 1/2%) per annum on the unpaid balance, both interest and principal being payable monthly as it accrues at the office of Mortgage Investment Corporation in Dallas, Texas in monthly installments of Twenty-five and 58/100 Dollars ($25.58) each, including interest, one on the first day of each month hereafter, commencing on the first day of April, 19 40, and continuing until the principal and interest are fully paid, except that the final payment of principal and interest, if not sooner paid, shall be due and payable on the first day of March, 19 65. The note also provides that if default is made in the payment of any installment thereunder, and if the default is not made good prior to the due date of the next such installment, at the option of the holder, the note shall become immediately due and payable without notice, and that this lien may be foreclosed. Failure to exercise this option is not to constitute a waiver of the right to exercise it in the event of any subsequent default. If the note is placed in the hands of an attorney for collection, or is collected through the Probate Court or the Bankruptcy Court or through other legal proceedings, the makers thereof agree to pay, as attorney's fees, an additional amount equal to ten per centum (10%) of the amount then owing on the note

The Grantors covenant as follows:

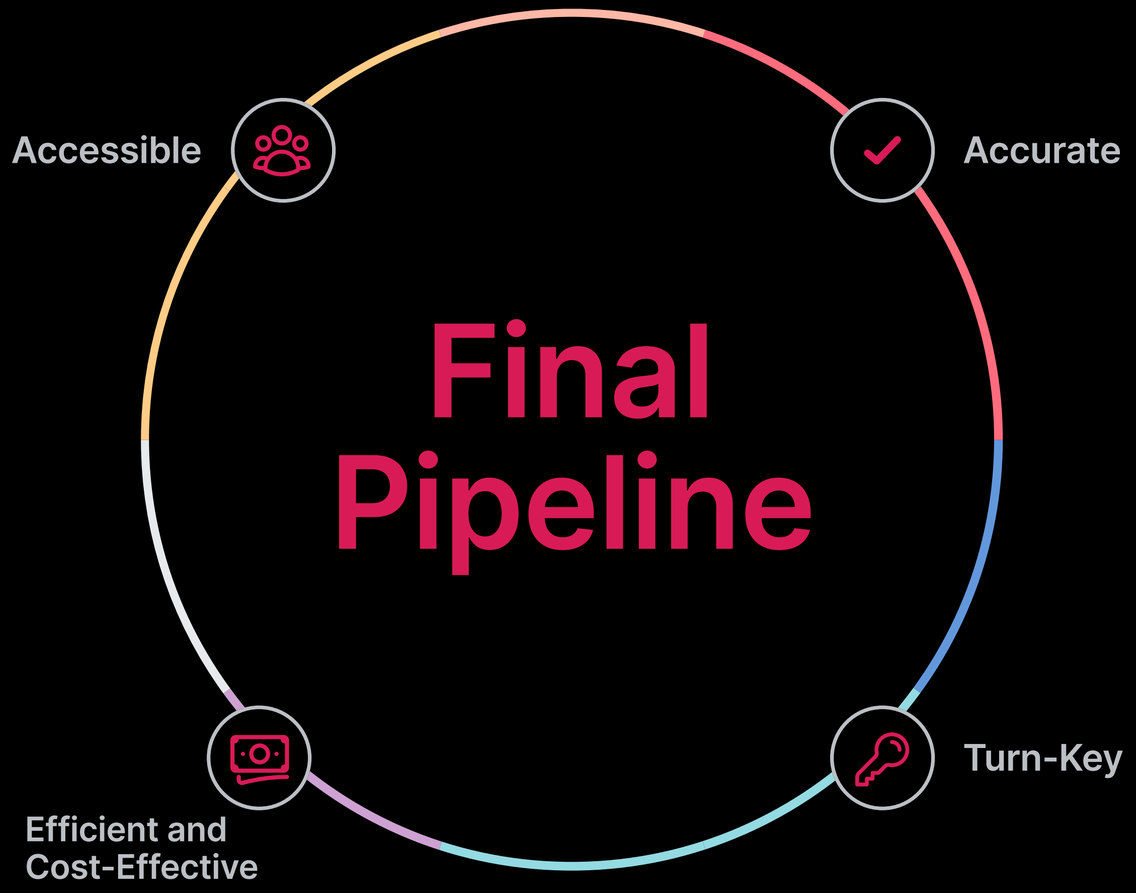
1. That they will pay the principal of and interest on the note secured hereby in accordance with the terms thereof. Privilege is reserved to pay the debt in whole, or in an amount equal to one or more monthly payments on the principal that are next due on the note, on the first day of any month prior to maturity; provided, however, that written notice of an intention to exercise such privilege is given at least thirty (30) days prior to prepayment; and provided further that in the event the debt is paid in full prior to maturity and at that time it is insured under the provisions of the National Housing Act, they will pay to the holder of the note an adjusted premium charge of one per centum (1%) of the original principal amount thereof, except that in no event shall the adjusted premium exceed the aggregate amount of premium charges which would have been payable if this Deed of Trust and the note secured hereby had continued to be insured until maturity, such payment to be applied by the holder of the note upon its obligation to the Federal Housing Administrator on account of mortgage insurance.

# CSV with 12 Features Per Doc

| E | F | G | H | I | J |
|---|---|---|---|---|---|
| Second_Borrower | Other_Party_First | Other_Party_Last | Lending_Bank | Interest_Rate | Loan_Amount |
| N/A | J. G. | HESTWOOD | FIRST MORTGAGE COMPANY OF HOUSTO | 4.50% | $3900.00 |
| N/A | J.G. | ESTWOOD | FIRST MORTGAGE COMPANY OF HOUSTO | 4.50% | $3550.00 |
| Berndt | Wm. S. | Bradley | Mortgage Investment Corporation | 4.50% | $4600.00 |
| N/A | R. M. | Orth | The Texas City National Bank | 4.50% | $3,000.00 |
| Quinn | B. B. | Yeager | Gulf Coast Investment Corporation | 4.50% | $4,400.00 |
| N/A | J.E. | Aosashe Jr. | J. E. FOSTER & SON, INC. | 4.50% | $3,950.00 |
| N/A | J. | Hestwood | First Mortgage Company of Houston, Inc. | 4.50% | $4,950.00 |

=

# Eventually the First Large Scale Public Dataset for FHA/VA Lending Research




The Objective

# Develop an accurate yet efficient, cost-effective, and turn-key approach for digitizing real-estate documents at scale.

- Current Manual Labelling Budget: $300,000
- Human Labeling Time Per Document: 5-10 Minutes

Existing Literature

YOLO APPROACH

# #1 Manually Enter Data

COMPUTER VISION APPROACH

# #2 Document Layout Analysis + OCR + Entity Recognition

## CV PIPELINE OVERVIEW

- Use CV Models to Identify Document Layout and Extract Areas of Interest
- Run OCR on Extracted Segments to get Text with Higher Accuracy
- Use SpaCY or LLMs for Extracting Entities from Segment Strings

**3**

**Computer Vision Models Tested from Meta's Detectron Library**

**2**

**OCR Models Tested**

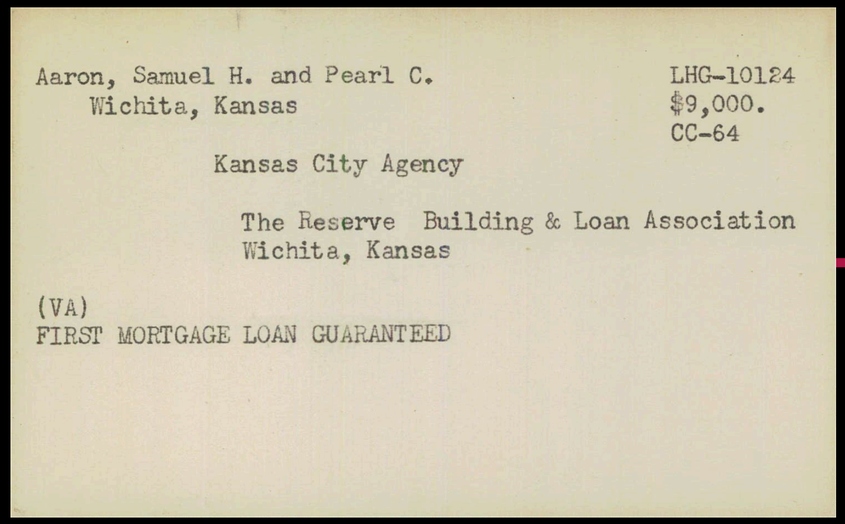
**3**

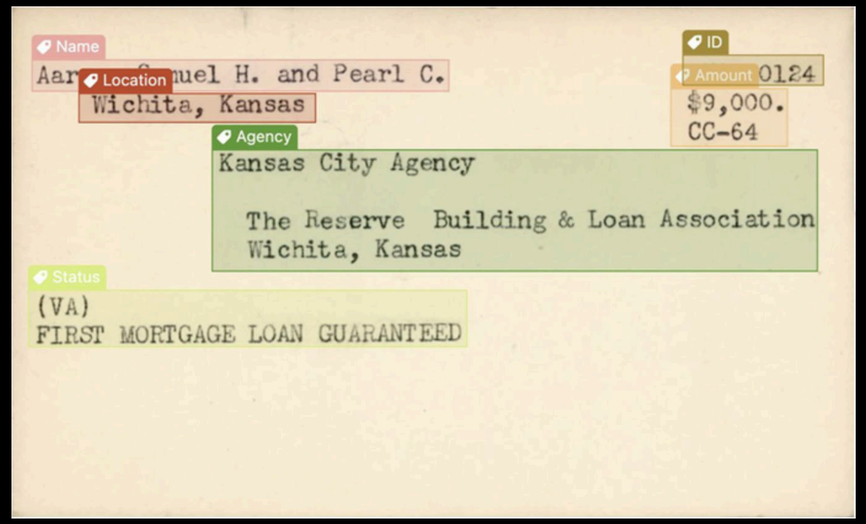
**Entity Recognition Models Tested including HuggingFace models and SpaCy**

# Why did they do this?

- OCR models (Tesseract/Adobe) were not robust enough to parse entire documents.
- Entity Recognition models were not powerful enough entirely handle downstream processing.

Detectron models output bounding boxes and entity label predictions. This solved both issues - segmented the document for better OCR accuracy and classified boxes with entity labels to reduce downstream processing.



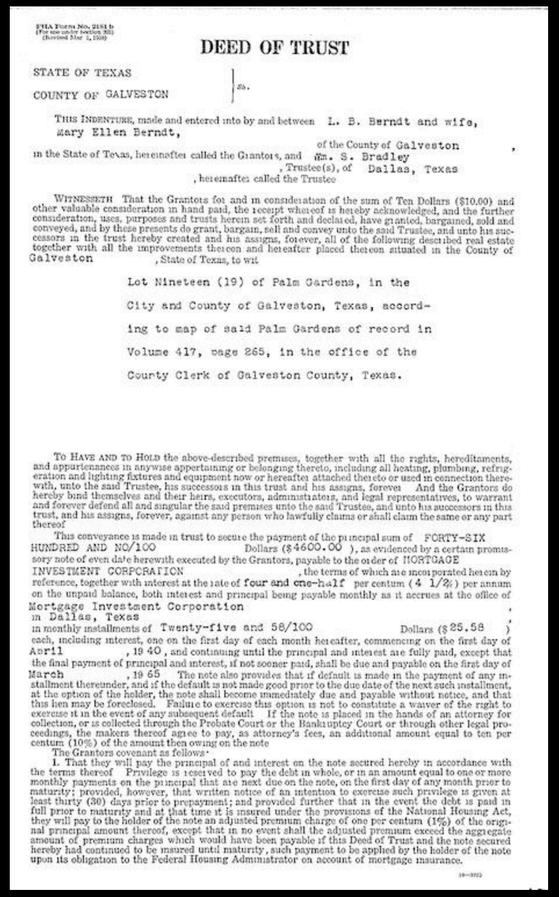


{
“Text”: “Kansas City Agency \n The Reserve Building & Loan Association Wichita, Kansas”,

“Label”: “Agency”
}

# Lack of Transferability

Below are document samples from Galveston county for the year 1940. As can be seen, **these documents exhibit high layout variability and textual density.**



FHA Form No. 2153 b
(For use under Section 203)
(Revised May 1, 1940)

DEED OF TRUST

STATE OF TEXAS
COUNTY OF GALVESTON } ss.

THIS INDENTURE, made and entered into by and between L. B. Berndt and wife, Mary Ellen Berndt, of the County of Galveston, in the State of Texas, hereinafter called the Grantors, and Wm. S. Bradley, Trustee(s), of Dallas, Texas, hereinafter called the Trustee

WITNESSETH That the Grantors for and in consideration of the sum of Ten Dollars ($10.00) and other valuable consideration in hand paid, the receipt whereof is hereby acknowledged, and the further consideration, uses, purposes and trusts herein set forth and declared, have granted, bargained, sold and conveyed, and by these presents do grant, bargain, sell and convey unto the said Trustee, and unto his successors in the trust hereby created and his assigns, forever, all of the following described real estate together with all the improvements thereon and hereafter placed thereon situated in the County of Galveston, State of Texas, to wit

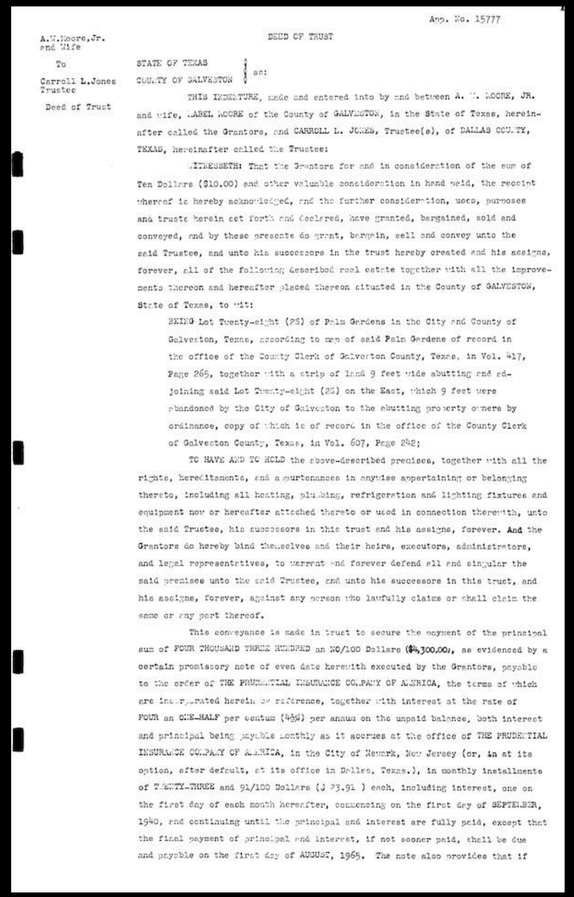
Lot Nineteen (19) of Palm Gardens, in the City and County of Galveston, Texas, according to map of said Palm Gardens of record in Volume 417, page 265, in the office of the County Clerk of Galveston County, Texas.

TO HAVE AND TO HOLD the above-described premises, together with all the rights, hereditaments, and appurtenances in anywise appertaining or belonging thereto, including all heating, plumbing, refrigeration and lighting fixtures and equipment now or hereafter attached thereto or used in connection therewith, unto the said Trustee, his successors in this trust and his assigns, forever. And the Grantors do hereby bind themselves and their heirs, executors, administrators, and legal representatives, to warrant and forever defend all and singular the said premises unto the said Trustee, and unto his successors in this trust, and his assigns, forever, against any person who lawfully claims or shall claim the same or any part thereof.

This conveyance is made in trust to secure the payment of the principal sum of FORTY-SIX HUNDRED AND NO/100 Dollars ($4600.00 ), as evidenced by a certain promissory note of even date herewith executed by the Grantors, payable to the order of MORTGAGE INVESTMENT CORPORATION, the terms of which are incorporated herein by reference, together with interest at the rate of four and one-half per centum ( 4 1/2%) per annum on the unpaid balance, both interest and principal being payable monthly as it accrues at the office of Mortgage Investment Corporation in Dallas, Texas in monthly installments of Twenty-five and 58/100 Dollars ($ 25.58 ) each, including interest, one on the first day of each month hereafter, commencing on the first day of April, 19 40 , and continuing until the principal and interest are fully paid, except that the final payment of principal and interest, if not sooner paid, shall be due and payable on the first day of March , 19 65 . The note also provides that if default is made in the payment of any installment thereunder, and if the default is not made good prior to the due date of the next such installment, at the option of the holder, the note shall become immediately due and payable without notice, and that this lien may be foreclosed. Failure to exercise this option is not to constitute a waiver of the right to exercise it in the event of any subsequent default. If the note is placed in the hands of an attorney for collection, or is collected through the Probate Court or the Bankruptcy Court or through other legal proceedings, the makers thereof agree to pay, as attorney's fees, an additional amount equal to ten per centum (10%) of the amount then owing on the note.

The Grantors covenant as follows:

1. That they will pay the principal of and interest on the note secured hereby in accordance with the terms thereof. Privilege is reserved to pay the debt in whole, or in an amount equal to one or more monthly payments on the principal that are next due on the note, on the first day of any month prior to maturity; provided, however, that written notice of an intention to exercise such privilege is given at least thirty (30) days prior to prepayment; and provided further that in the event the debt is paid in full prior to maturity and at that time it is insured under the provisions of the National Housing Act, they will pay to the holder of the note an adjusted premium charge of one per centum (1%) of the original principal amount thereof, except that in no event shall the adjusted premium exceed the aggregate amount of premium charges which would have been payable if this Deed of Trust and the note secured hereby had continued to be insured until maturity, such payment to be applied by the holder of the note upon its obligation to the Federal Housing Administrator on account of mortgage insurance.



Acc. No. 15777

DEED OF TRUST

A.W.Moore,Jr. and Wife
To
Carroll L.Jones Trustee
Deed of Trust

STATE OF TEXAS
COUNTY OF GALVESTON } ss:

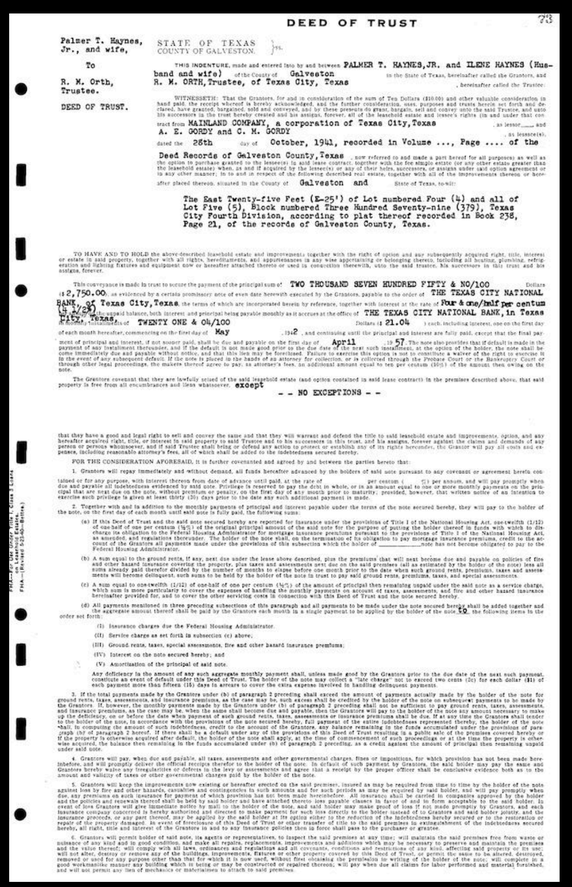
THIS INDENTURE, made and entered into by and between A. W. MOORE, JR. and wife, ARIEL MOORE of the County of GALVESTON, in the State of Texas, hereinafter called the Grantors, and CARROLL L. JONES, Trustee(s), of DALLAS COUNTY, TEXAS, hereinafter called the Trustee:

WITNESSETH: That the Grantors for and in consideration of the sum of Ten Dollars ($10.00) and other valuable consideration in hand paid, the receipt whereof is hereby acknowledged, and the further consideration, uses, purposes and trusts herein set forth and declared, have granted, bargained, sold and conveyed, and by these presents do grant, bargain, sell and convey unto the said Trustee, and unto his successors in the trust hereby created and his assigns, forever, all of the following described real estate together with all the improvements thereon and hereafter placed thereon situated in the County of GALVESTON, State of Texas, to wit:

BEING Lot Twenty-eight (28) of Palm Gardens in the City and County of Galveston, Texas, according to map of said Palm Gardens of record in the office of the County Clerk of Galveston County, Texas, in Vol. 417, Page 265, together with a strip of land 9 feet wide abutting and adjoining said Lot Twenty-eight (28) on the East, which 9 feet were abandoned by the City of Galveston to the abutting property owners by ordinance, copy of which is of record in the office of the County Clerk of Galveston County, Texas, in Vol. 607, Page 242;

TO HAVE AND TO HOLD the above-described premises, together with all the rights, hereditaments, and appurtenances in anywise appertaining or belonging thereto, including all heating, plumbing, refrigeration and lighting fixtures and equipment now or hereafter attached thereto or used in connection therewith, unto the said Trustee, his successors in this trust and his assigns, forever. And the Grantors do hereby bind themselves, their heirs, executors, administrators, and legal representatives, to warrant and forever defend all and singular the said premises unto the said Trustee, and unto his successors in this trust, and his assigns, forever, against any person who lawfully claims or shall claim the same or any part thereof.

This conveyance is made in trust to secure the payment of the principal sum of FOUR THOUSAND THREE HUNDRED an NO/100 Dollars ($4,300.00, as evidenced by a certain promissory note of even date herewith executed by the Grantors, payable to the order of THE PRUDENTIAL INSURANCE COMPANY OF AMERICA, the terms of which are incorporated herein by reference, together with interest at the rate of FOUR an ONE-HALF per centum (4½%) per annum on the unpaid balance, both interest and principal being payable monthly as it accrues at the office of THE PRUDENTIAL INSURANCE COMPANY OF AMERICA, in the City of Newark, New Jersey (or, in at its option, after default, at its office in Dallas, Texas.), in monthly installments of TWENTY-THREE and 91/100 Dollars ($ 23.91 ) each, including interest, one on the first day of each month hereafter, commencing on the first day of SEPTEMBER, 1940, and continuing until the principal and interest are fully paid, except that the final payment of principal and interest, if not sooner paid, shall be due and payable on the first day of AUGUST, 1965. The note also provides that if



DEED OF TRUST 73

Palmer T. Haynes, Jr., and wife,
To
R. M. Orth, Trustee.
DEED OF TRUST.

STATE OF TEXAS
COUNTY OF GALVESTON } ss.

THIS INDENTURE, made and entered into by and between PALMER T. HAYNES,JR. and ELENE HAYNES (Husband and wife) of the County of Galveston in the State of Texas, hereinafter called the Grantors, and R. M. ORTH, Trustee, of Texas City, Texas, hereinafter called the Trustee:

WITNESSETH: That the Grantors, for and in consideration of the sum of Ten Dollars ($10.00) and other valuable consideration in hand paid, the receipt whereof is hereby acknowledged, and the further considerations, uses, purposes and trusts herein set forth and declared, have granted, bargained, sold and conveyed, and by these presents do grant, bargain, sell and convey unto the said Trustee, and unto his successors in the trust hereby created and his assigns, forever, all of the hereinafter described real estate, together with all of the improvements thereon, subject however to that certain contract from MAINLAND COMPANY, a corporation of Texas City, Texas, as lessor, and A. E. DOWDY and C. N. DOWDY, as lessees, dated the 26th day of October, 1941, recorded in Volume ..., Page ..., of the Deed Records of Galveston County, Texas, now referred to and made a part hereof for all purposes; as well as the option to purchase granted to the lessees in said lease contract, together with the fee simple estate (or any other estate greater than said lease) in and under said lease, and the rights of the lessees, their heirs, successors and assigns, under said lease and option agreement or in any other manner, to do and in respect of the following described real estate, together with all of the improvements thereon, or hereafter placed thereon, situated in the County of Galveston and State of Texas, to-wit:

The East Twenty-five Feet (E-25') of Lot numbered Four (4) and all of Lot Five (5), Block numbered Three Hundred Seventy-nine (379), Texas City Fourth Division, according to plat thereof recorded in Book 238, Page 21, of the records of Galveston County, Texas.

TO HAVE AND TO HOLD the above-described leasehold estate and improvements, together with the right of option and any subsequently acquired right, title, interest or estate in said property, together with all rights, hereditaments, and appurtenances in any wise appertaining or belonging thereto, including all heating, plumbing, refrigeration and lighting fixtures and equipment now or hereafter attached thereto or used in connection therewith, unto the said trustee, his successors in this trust and his assigns, forever.

This conveyance is made in trust to secure the payments of the principal sum of TWO THOUSAND SEVEN HUNDRED FIFTY & NO/100 Dollars ($2,750.00, as evidenced by a certain promissory note of even date herewith executed by the Grantors, payable to the order of THE TEXAS CITY NATIONAL BANK, Texas City, Texas, the terms of which are incorporated herein by reference, together with interest at the rate of Four & one/half per centum (4 1/2%) per annum on the unpaid balance, both interest and principal being payable monthly as it accrues at the office of THE TEXAS CITY NATIONAL BANK, in Texas City, Texas, in monthly installments of TWENTY ONE & 04/100 Dollars ($ 21.04 ) each, including interest, one on the first day of each month hereafter, commencing on the first day of May , 19 42 , and continuing until the principal and interest are fully paid, except that the final payment of principal and interest, if not sooner paid, shall be due and payable on the first day of April , 19 57 . The note also provides that if default is made in the payment of any installment thereunder, and if the default is not made good prior to the due date of the next such installment, at the option of the holder, the note shall become immediately due and payable without notice, and that this lien may be foreclosed. Failure to exercise this option is not to constitute a waiver of the right to exercise it in the event of any subsequent default. If the note is placed in the hands of an attorney for collection, or is collected through the Probate Court or the Bankruptcy Court or through other legal proceedings, the makers thereof agree to pay, as attorney's fees, an additional amount equal to ten per centum (10%) of the amount then owing on the note.

The Grantors covenant that they are lawfully seized of the said leasehold estate (and option contained in said lease contract) in the premises described above, that said property is free from all encumbrances and liens whatsoever, except

- - NO EXCEPTIONS - -

that they have a good and legal right to sell and convey the same and that they will warrant and defend the title to said leasehold estate and improvements, option, and any hereafter acquired right, title, or interest in said property to said Trustee and to his successors in this trust, and his assigns, forever against the claims and demands of any person or persons whomsoever, and if said Trustee shall bring or defend any action in connection with this trust, and his assigns, forever against the claims and demands of any person or persons whomsoever, and if said Trustee shall bring or defend any action in connection with this trust, and his assigns, forever against the claims and demands of any person or persons whomsoever, and if said Trustee shall bring or defend any action in connection with this trust, or become involved in any litigation on account of its ownership hereunder, the Grantor will pay all costs and expenses, including reasonable attorney's fees, all of which shall be added to the indebtedness secured hereby.

FOR THE CONSIDERATION AFORESAID, it is further covenanted and agreed by and between the parties hereto that:

1. Grantors will repay immediately and without demand, all funds hereafter advanced by the holders of said note pursuant to any covenant or agreement herein contained or for any purpose, with interest thereon from date of advance until paid, at the rate of ... per centum ( ...%) per annum, and will pay promptly when due and payable all indebtedness evidenced by said note. Privilege is reserved to pay the debt in whole, or in an amount equal to one or more monthly payments on the principal that are next due on the note, without premium or penalty, on the first day of any month prior to maturity; provided, however, that written notice of an intention to exercise such privilege is given at least thirty (30) days prior to the date any such additional payment is made.

2. Together with and in addition to the monthly payments of principal and interest payable under the terms of the note secured hereby, they will pay to the holder of the note, on the first day of each month until said note is fully paid, the following sums:

(a) If this Deed of Trust and the said note secured hereby are reported for insurance under the provisions of Title I of the National Housing Act, one-twelfth (1/12) of one-half of one per centum (½%) of the original principal amount of the said note for the purpose of putting the holder thereof in funds with which to discharge the obligation of the holder hereof to pay the premium charges payable pursuant to the provisions of Title I of the National Housing Act, as amended, and regulations thereunder. The holder of the note shall, on the termination of its obligation to pay mortgage insurance premiums, credit to the account of the Grantors all payments made under the provisions of this subsection with the holder of ..... note has not become obligated to pay to the Federal Housing Administrator.

(b) A sum equal to the ground rents, if any, next due under the lease above described, plus the premiums that will next become due and payable on policies of fire and other hazard insurance covering the property, plus taxes and assessments next due on the said premises (all as estimated by the holder of the note) less all sums already paid therefor divided by the number of months to elapse before one month prior to the date when such ground rents, premiums, taxes and assessments will become delinquent, such sums to be held by the holder of the note in trust to pay said ground rents, premiums, taxes, and special assessments.

(c) A sum equal to one-twelfth (1/12) of one-half of one per centum (½%) of the amount of principal then remaining unpaid under the said note as a service charge, which sum is more particularly to cover the expenses of handling the monthly payments on account of taxes, assessments, and fire and other hazard insurance hereinafter provided for, and to cover the other servicing costs in connection with this Deed of Trust and the note secured hereby.

(d) All payments mentioned in these preceding subsections of this paragraph and all payments to be made under the note secured hereby shall be added together and the aggregate amount thereof shall be paid by the Grantors each month in a single payment to be applied by the holder of the note to the following items in the order set forth:

(I) Insurance charges due the Federal Housing Administrator.
(II) Service charges as set forth in subsection (c) above.
(III) Ground rents, taxes, special assessments, fire and other hazard insurance premiums.
(IV) Interest on the note secured hereby, and
(V) Amortization of the principal of said note.

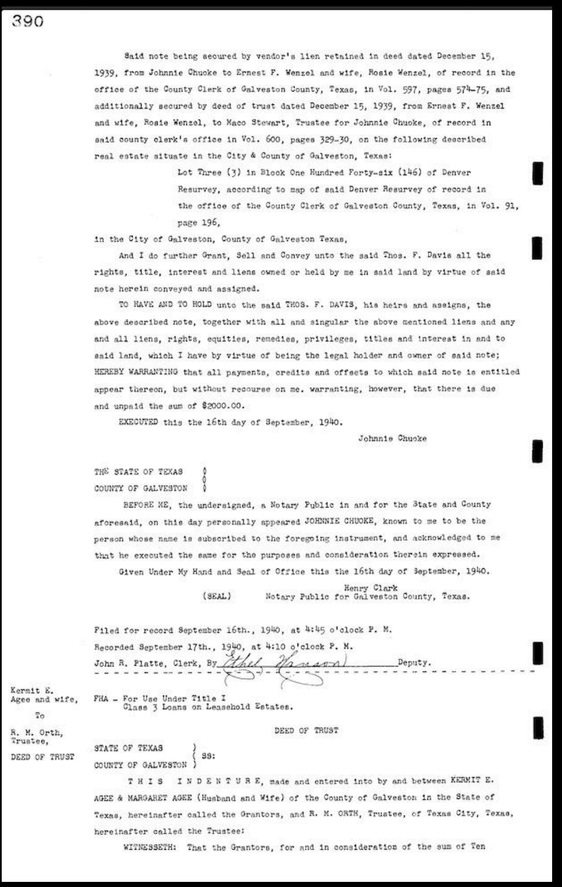
Any deficiency in the amount of any such aggregate monthly payment shall, unless made good by the Grantors prior to the due date of the next such payment, constitute an event of default under this Deed of Trust. The holder of the note may collect a "late charge" not to exceed two cents (2c) for each dollar ($1) of each payment more than fifteen (15) days in arrears to cover the extra expense involved in handling delinquent payments.

3. If the total payments made by the Grantors under (b) of paragraph 2 preceding shall exceed the amount of payments actually made by the holder of the note for ground rents, taxes or assessments, or insurance premiums, as the case may be, such excess shall be credited by the holder of the note on subsequent payments to be made by the Grantors. If, however, the monthly payments made by the Grantors under (b) of paragraph 2 preceding shall not be sufficient to pay ground rents, taxes, assessments, and insurance premiums, as the case may be, when the same shall become due and payable, then the Grantors will pay to the holder of the note any amount necessary to make up the deficiency, on or before the date when payment of such ground rents, taxes, assessments or insurance premiums shall be due. If at any time the Grantors shall tender to the holder of the note, in accordance with the provisions of the note secured hereby, full payment of the entire indebtedness represented thereby, the holder of the note shall, in computing the amount of such indebtedness, credit to the account of the Grantors any balance remaining in the funds accumulated under the provisions of clause (b) of paragraph 2 hereof. If there shall be a default under any of the provisions of this Deed of Trust resulting in a public sale of the premises covered hereby or if the property is otherwise acquired after default, the holder of the note shall apply, at the time of commencement of such proceedings or at the time the property is otherwise acquired, the balance then remaining in the funds accumulated under (b) of paragraph 2 preceding, as a credit against the amount of principal then remaining unpaid under said note.

4. Grantors will pay, when due and payable, all taxes, assessments and other governmental charges, fines or impositions, for which provision has not been made hereinbefore, and will promptly deliver the official receipts therefor to the holder of the note. In default of such payment by Grantors, the said holder may pay the same and may make any tax search or obtain any tax certificate which in its judgment is necessary to protect the lien hereof, and in any such case the said holder shall be the sole judge of the legality or the proper officer shall be conclusive evidence both as to the amount and validity of taxes or other governmental charges paid by the holder of the note.

5. Grantors will keep the improvements now existing or hereafter erected on the said premises, insured as may be required from time to time by the holder of the note against loss by fire and other hazards, casualties and contingencies in such amounts and for such periods as may be required by said holder, and will pay promptly when due any premiums on such insurance for payment of which provision has not been made hereinbefore. All insurance shall be carried in companies approved by the holder and the policies and renewals thereof shall be held by said holder and have attached thereto loss payable clauses in favor of and in form acceptable to the said holder. In event of loss Grantors will give immediate notice by mail to the holder of the note, and said holder may make proof of loss if not made promptly by Grantors, and each insurance company concerned is hereby authorized and directed to make payment for such loss directly to the said holder instead of to the Grantors and Grantors' insurance proceeds, or any part thereof, may be applied by the said holder at its option either to the reduction of the indebtedness hereby secured or to the restoration or repair of the property damaged. In event of foreclosure of this Deed of Trust or other transfer of title to the said premises in extinguishment of the indebtedness secured hereby, all right, title and interest of the Grantors in and to any insurance policies then in force shall pass to the purchaser or grantee.

6. Grantors will permit holder of said note, its agents or representatives, to inspect the said premises at any time; will maintain the said premises free from waste or nuisance of any kind and in good condition, and make all repairs, replacements, improvements and additions which may be necessary to preserve and maintain the premises and the value thereof; will comply with all laws, ordinances and regulations and all covenants, conditions and restrictions of any kind, affecting said property or its use; will not remove or demolish any building thereon; will not commit or permit any waste thereof; will not make any alterations in the premises without the consent of the holder of the note; will not remove or used for any purpose other than that for which it is now used, without first obtaining the permission in writing of the holder of the note; will complete in a good workmanlike manner any building which is being or may be constructed or repaired thereon; will pay when due all claims for labor performed and material furnished, and will not permit any lien of mechanics or materialmen to attach to said premises.



390

Said note being secured by vendor's lien retained in deed dated December 15, 1939, from Johnnie Chuoke to Ernest F. Wenzel and wife, Rosie Wenzel, of record in the office of the County Clerk of Galveston County, Texas, in Vol. 597, pages 574-75, and additionally secured by deed of trust dated December 15, 1939, from Ernest F. Wenzel and wife, Rosie Wenzel, to Mace Stewart, Trustee for Johnnie Chuoke, of record in said county clerk's office in Vol. 600, pages 329-30, on the following described real estate situate in the City & County of Galveston, Texas:

Lot Three (3) in Block One Hundred Forty-six (146) of Denver Resurvey, according to map of said Denver Resurvey of record in the office of the County Clerk of Galveston County, Texas, in Vol. 91, page 196,

in the City of Galveston, County of Galveston Texas,

And I do further Grant, Sell and Convey unto the said Thos. F. Davis all the rights, title, interest and liens owned or held by me in said land by virtue of said note herein conveyed and assigned.

TO HAVE AND TO HOLD unto the said THOS. F. DAVIS, his heirs and assigns, the above described note, together with all and singular the above mentioned liens and any and all liens, rights, equities, remedies, privileges, titles and interest in and to said land, whilst I have by virtue of being the legal holder and owner of said note; HEREBY WARRANTING that all payments, credits and offsets to which said note is entitled appear thereon, but without recourse on me, warranting, however, that there is due and unpaid the sum of $2000.00.

EXECUTED this the 16th day of September, 1940.

Johnnie Chuoke

THE STATE OF TEXAS }
COUNTY OF GALVESTON }

BEFORE ME, the undersigned, a Notary Public in and for the State and County aforesaid, on this day personally appeared JOHNNIE CHUOKE, known to me to be the person whose name is subscribed to the foregoing instrument, and acknowledged to me that he executed the same for the purposes and consideration therein expressed.

Given Under My Hand and Seal of Office this the 16th day of September, 1940.

(SEAL) Henry Clark
Notary Public for Galveston County, Texas.

Filed for record September 16th., 1940, at 4:45 o'clock P. M.
Recorded September 17th., 1940, at 4:10 o'clock P. M.
John R. Platte, Clerk, By ... Deputy.

Kermit E. Agee and wife, Margaret Agee
To
R. M. Orth, Trustee,
DEED OF TRUST

FHA - For Use Under Title I
Class 3 Loans on Leasehold Estates.

DEED OF TRUST

STATE OF TEXAS }
COUNTY OF GALVESTON } ss:

THIS INDENTURE, made and entered into by and between KERMIT E. AGEE & MARGARET AGEE (Husband and Wife) of the County of Galveston in the State of Texas, hereinafter called the Grantors, and R. M. ORTH, Trustee, of Texas City, Texas, hereinafter called the Trustee:

WITNESSETH: That the Grantors, for and in consideration of the sum of Ten

# Renewed Approach

1

## Modern OCR Models

Modern open and closed-source CNN & RNN based OCR models exist that excel at word recognition.

2

## Parse Entire Documents

Parse entire documents eliminating the CV step for accuracy.

PyTorch Models, therefore GPU Acceleration.

3

## Introduce LLMs

LLMs understand context.

They can also smooth and preprocess OCR errors.

Prompt-Engineering will help us tune LLMs for out extraction task.

4

## Validate Output and Fine-Tune

Test different OCR models and LLMs.

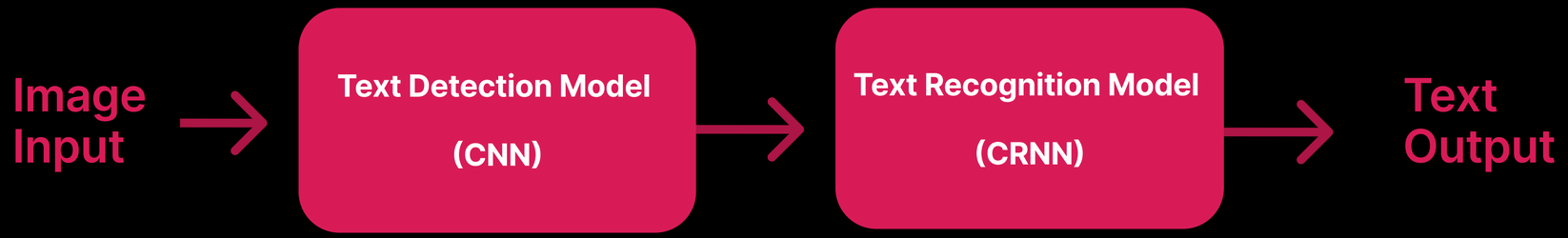
Validate with relevant metrics.

Analyze errors to find ideal combinations.

OCR Stage

LEVERAGING MODERN CRNN-BASED OCR MODELS

# 2-Step OCR Workflow




## Library Choice

**docTR**

- Higher Accuracy
- Model Choices
- Better API

## Model Choice

Text-Detection Model: **DB Resnet 50**

Text-Recognition Model: **CRNN VGG 16**
(VGG 16+Bi-Directional RNN)

Closest to Amazon Textract

**Inference Speed: 19 Seconds per Doc (L4 GPU)**

## OCR MODELS TESTED

docTR

Easy OCR

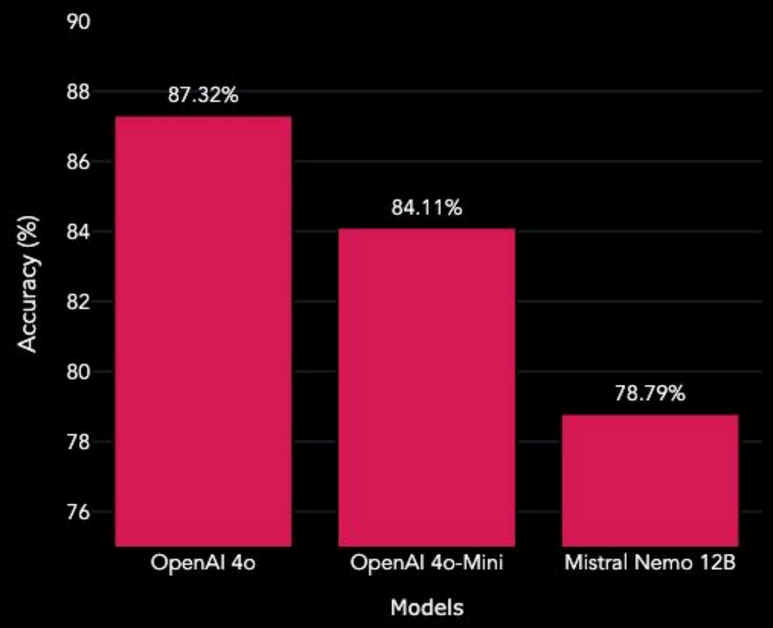
Tesseract OCR

LLM Stage

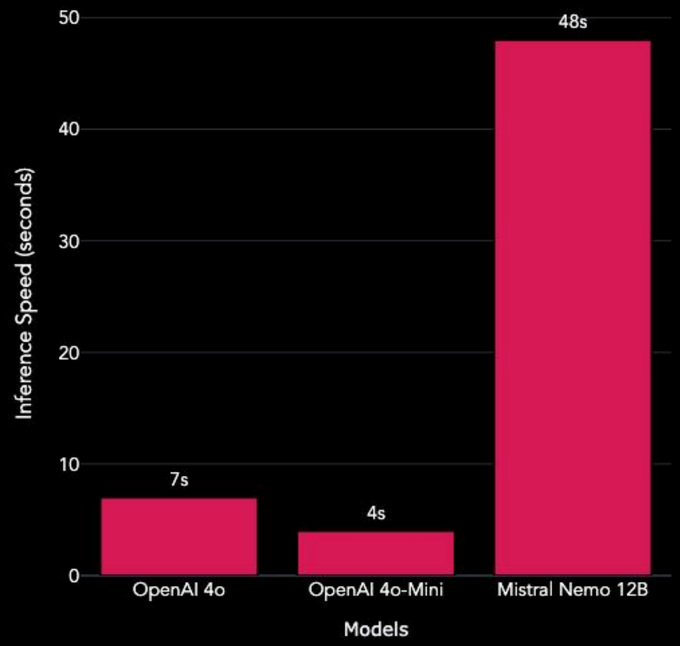
TESTING LARGE LANGUAGE MODELS

# Open-Source vs Gated Models

## ACCURACY




## SPEED




*2 V100 GPUs for Open-Source Models

Open-Source models are currently not on-par with OpenAI models for our task.

LLMS TESTED

OpenAI

Meta

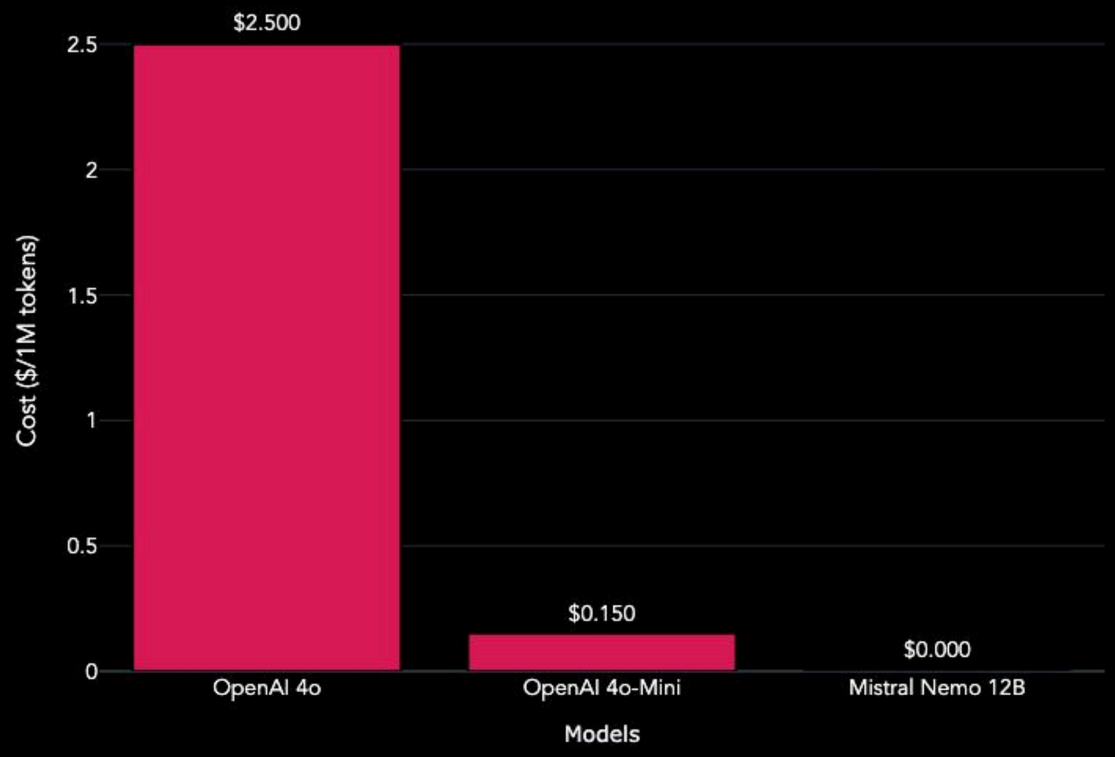
MISTRAL AI_

NVIDIA

LLM Stage

TESTING LARGE LANGUAGE MODELS

# Open-Source vs Gated Models




As an initial start, we believe the cost is justified.
However, as we scale, we must consider otherwise.

LLMS TESTED

OpenAI

Meta

MISTRAL AI_

NVIDIA

Model Validation

ACCURACY IN THE STRINGS WORLD

# Levenshtein Distance

**Counts the minimum number of single-character edits (insertions, deletions, or substitutions) needed to transform one string into the other.**

**We used Tokenized Levenshtein Distance.**

1. Tokenize String - Ignores Weird Punctuation
2. Sort them Alphabetically - Ignores Order
3. Calculate Levenshtein Distance
4. Normalize by String Length - Get % Accuracy

**Accuracy("Let's cook Grandma", "Let's cook, Grandma") = 100%**

**Insights taken from Model Validation were used to refine prompts.**

## IMPACT OF PROMPT ENGINEERING

**72% Initial**
with Original Prompt

**87% Final**
Prompt Changes from Validation Analysis

**+15% Gain**
More Potential.....

Final Pipeline

2 Page Input .TIF
xN

JSON Output

docTR OCR Stage

(Resnet + VGG RCNN)

JSON Output

LLM Stage

OpenAI 4o

Sanity Checks

.CSV with Entity Extractions
Nx12

By The Numbers

# Final Approach

**87 %**

Accuracy

Levenshtein Distance

**26 s**

Per Document

OCR + API Call

**.006 $**

Per Document

Average Doc is 2500 tokens

**1773 $**

For Entire Corpus

280,000 Documents

Next Steps

# The Future

### Scale

Scale to thousands of documents. Will understand more model nuances and refine prompts accordingly.

### Go Open-Source

Llama 3.3 just pulled up. Put that to the test.

### Publish

Goal has been to make the dataset and code public. So, once internal testing is done, publish.

Thanks.

Thanks.
Thanks.
Thanks.
Thanks.
Thanks.