



medsync



**Weill Cornell
Medicine**

**CORNELL
TECH**

HOME OF THE
**JACOBS
INSTITUTE**



How Might We: Transcribe handwritten medical charts into electronic searchable text?

Potential Customer Testimonials

“I would absolutely use this product; in working for the [Department of Health], being able to increase specification of historical vaccine logs would be critical in **today's climate of disease recurrence**.”

“While medical research is [nearly] endless, epidemiologic studies get scarce prior to ~100 years ago. In **increasing data accessibility**, we would be better prepared to run meaningful, comprehensive studies.”

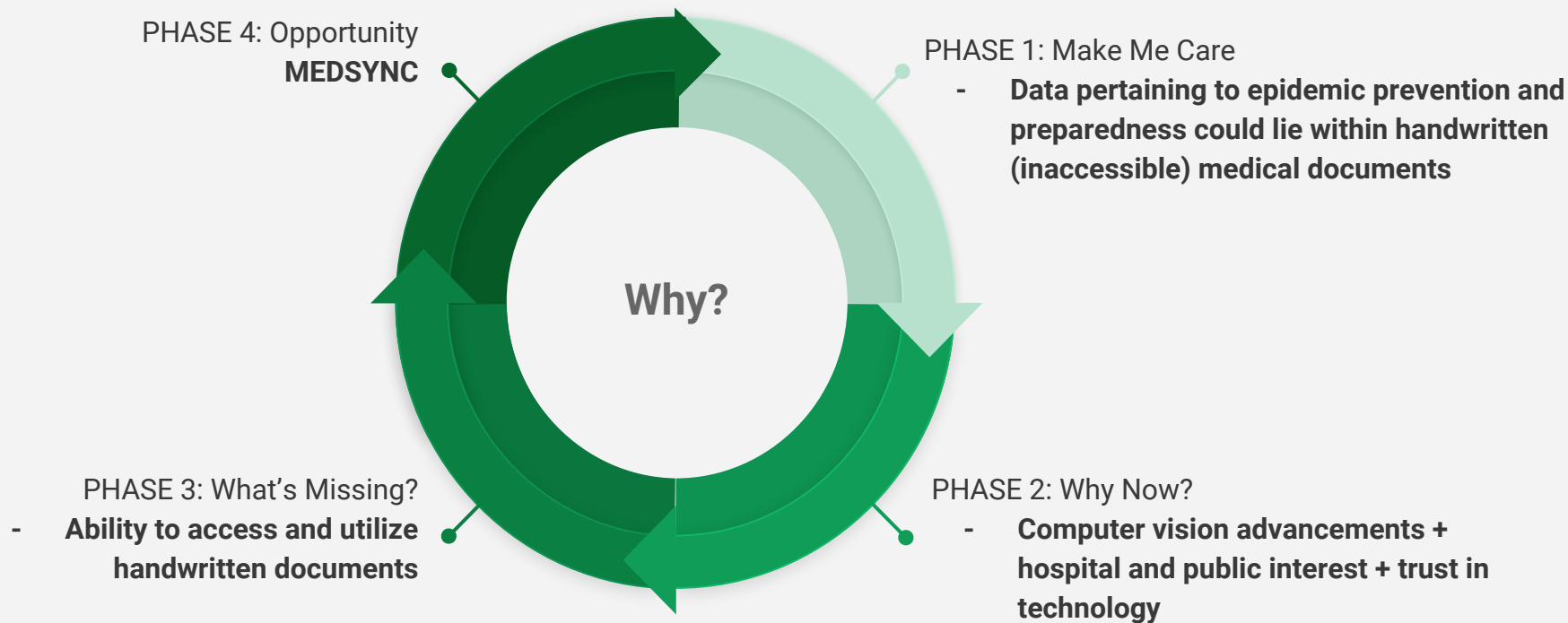
“Epidemiologic data from the 18th and 19th century is literally blank, and making them available can greatly advance our research on the **risk factors of diseases**.”

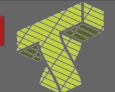
“The lack of historical dataset is one of the biggest obstacles to epidemiology research, and the **auto-generation of dataset from historical medical charts** will save tremendous amount of time and efforts for public health researchers”

“**Information** is the lifeblood of medicine”

- Vast archives at New York Presbyterian
- 142 teaching hospitals in the US, each potentially having own unique libraries
- Over 400,000 pieces of printed work at the National Library of Medicine







medsync

Has been in America four years - eighteen months ago began to suffer from a severe pain commencing in the left hip and thence extending down the posterior and outer aspect of the leg to the ankle - this pain has been constant ever since - it is more severe in damp and unpleasant weather - aggravated by motion and relieved by rest - general health good - appetite poor - menstruates regularly - bowels rather confined - pulse natural - skin cool

Aug 23rd Ordered Electricity by acupuncture

Patient has complained of more pain since the application of electricity - ordered to be discontinued

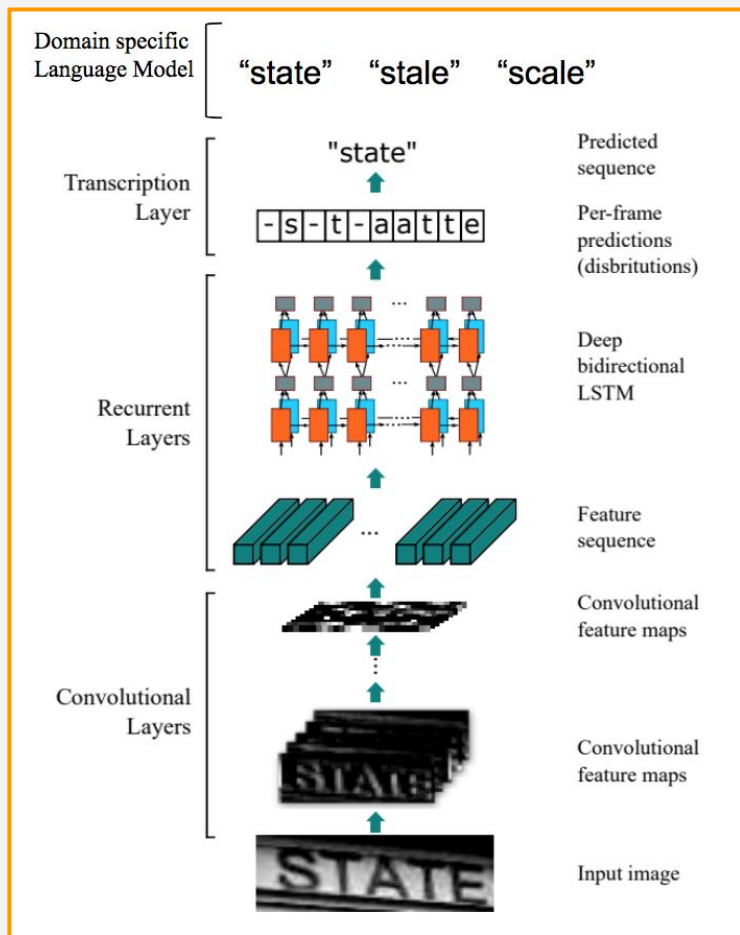
Ordered Amp. vesic. along course of sciatic nerve -

Aug 28th Patient has somewhat improved pain entirely gone - since last note patient has had five or six blisters applied

15th S. Cured

Constipation

960



Exact Word Match Accuracy

Vision GCP API	Language Model Output
~8%	73.90%

Next Step:

↑ Data == ↑ Accuracy

1. Unclear between mop, hop and top
2. Worse for longer words



Sprint 1

- Lack of focused market/
target audience
- Minimal training data

Sprint 2

- Focused market and audience
- New training data
- Lack of focused user story

Sprint 3

- Focused market and
audience
- Better training data and test
data
- Focused user story

Sprint 1

- Lack of focused market/
target audience
- Minimal training data

Sprint 2

- Focused market and audience
- New training data
- Lack of focused user story

Sprint 3

- Focused market and
audience
- Better training data and test
data
- Focused user story

Sprint 1

- Lack of focused market/target audience
- Minimal training data

Sprint 2

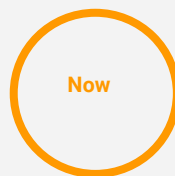
- Focused market and audience
- New training data
- Lack of focused user story

Sprint 3

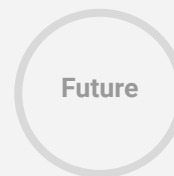
- Focused market and audience
- Better training data and test data
- Focused user story



Before



Now



Future

Transcribathons

Limited amounts of training data

Working ML Model

Works primarily with training data
from online sources

Create end-to-end model that is
capable of transcribing NYP
archives

Need to box words

Increase training data from correct
century to maximize accuracy