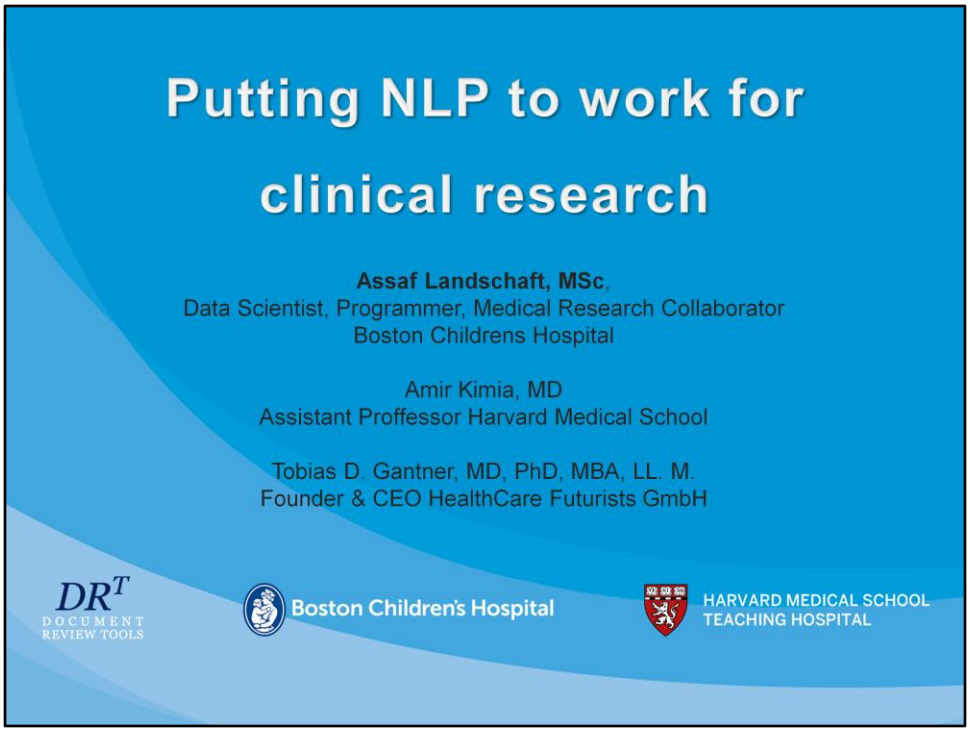


Hello I am Assaf.

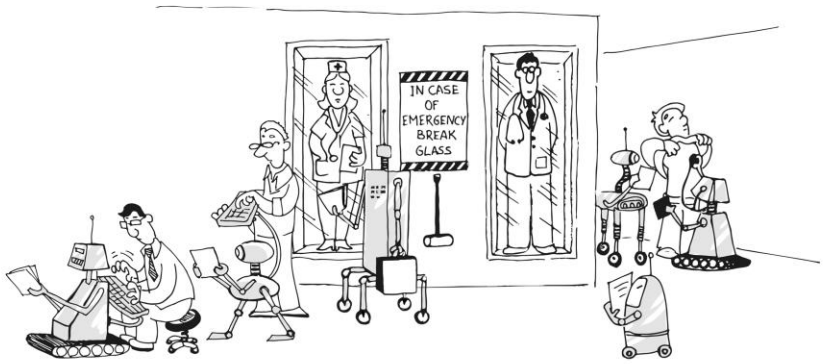
Around seven years ago I had to write a masters project on Active Learning.

I wanted to work with real data and solve a real problem. I ended up with Amir Kimia from BCH and HMS.
Since then, my collaboration with Amir's group contributed to more than 50 peer reviewed publications, including in the NeJM.

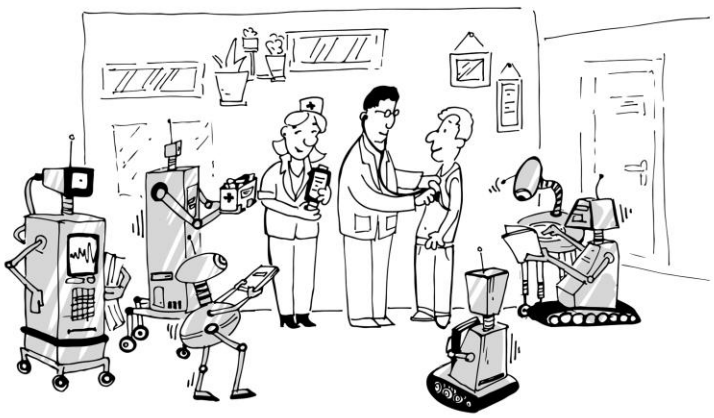
Since I live in Bergisch Gladbach near Cologne, I know Tobias Gantner from HealthCare Futurists and he invited me to this conference to talk about our work.

But before I start, I would like to show you our vision on the topic of this conference – The Future of Healthcare

NOT our vision



Our vision



In our visions the clinicians stand at the center.

The interaction between clinicians and computers is direct. There are no technical people.

The inspiration for our vision is the basic change that AI has brought to the IT world.

I would like to explain this change:

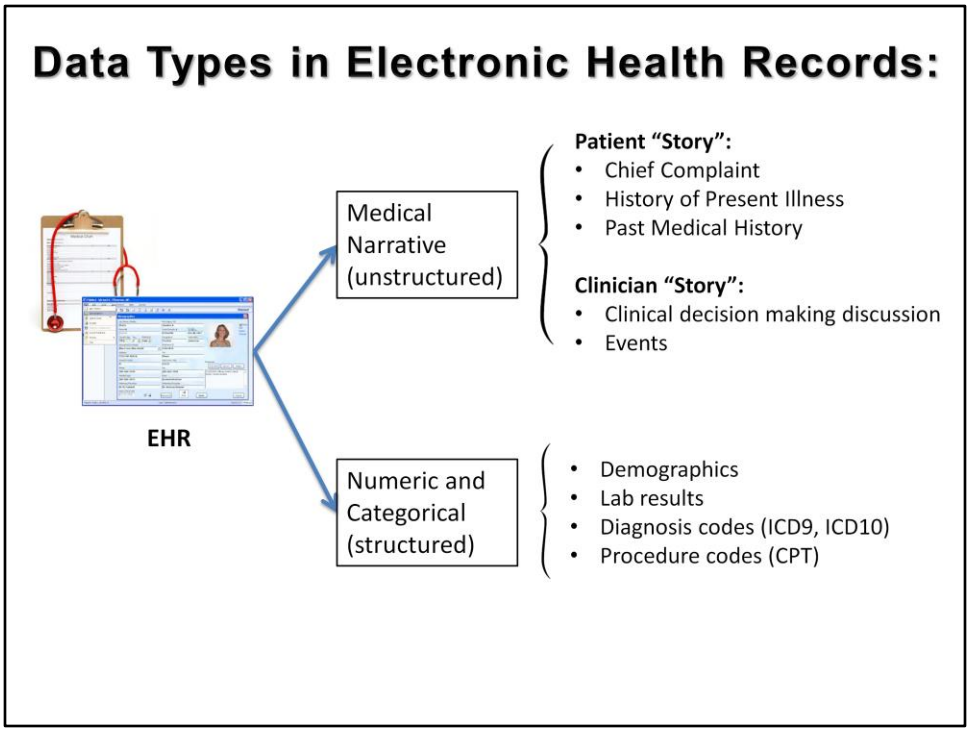
AI vs. Traditional Programming

- In traditional programming we build the descision making rules directly into the program using programming languages
- In AI we build an agent that learns the descision making rules from data examples

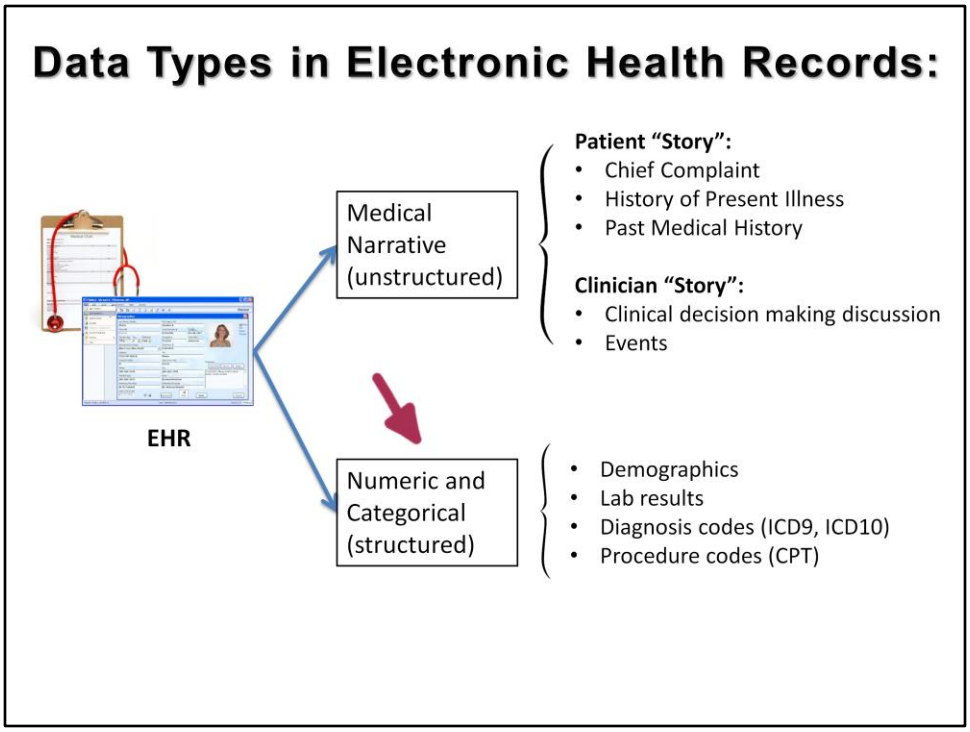
Strep Throat infection: **(**bacterial infection**)**
1. Fever
2. Swollen lymph nodes
3. No cough

...

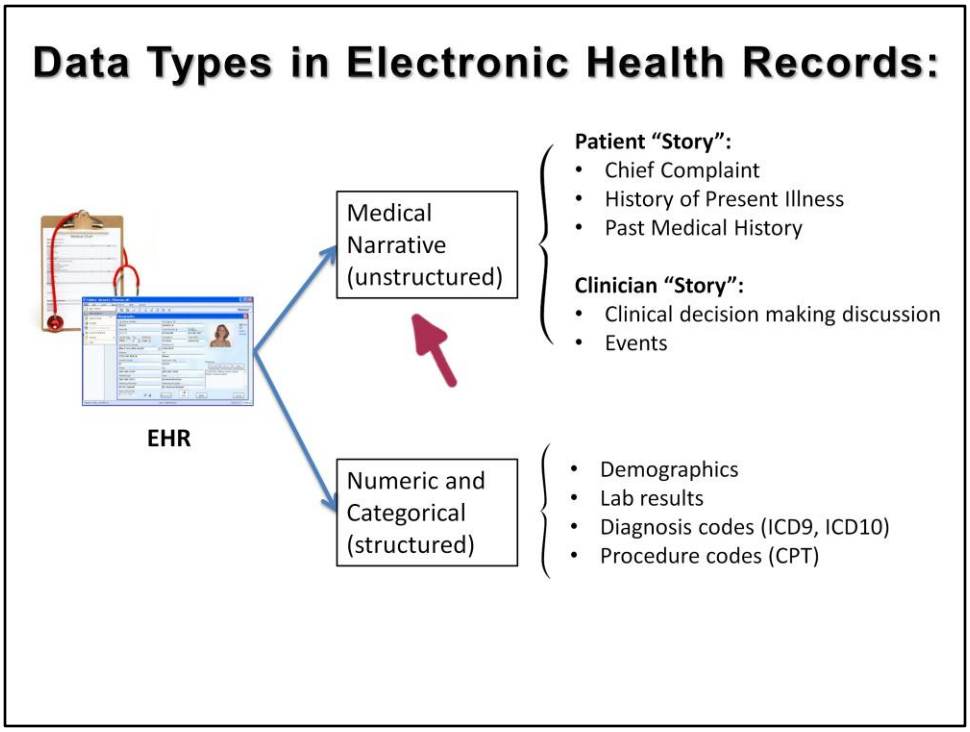
So, we have started to talk about medical data, let‘s see how do clinical data look like?



There are 2 types of data that we find in electronic health records:**<click>**



Structured data, for example Demographics: (address, age, height, weight) **<click>**



And unstructured data that we like to call **Medical Narrative; These are the Stories** of the patient and the clinician<click>

Knee:

Pain or discomfort, in the morning, climbing stairs, standing, driving

Breastfeeding Counselors.

Issues: Pain as the baby latches, Minutes until the baby **falls asleep at the breast, hurting nipples between feedings**

The truth is in the narrative

- Most medical data exist exclusively in the free text format of the clinical notes, despite efforts to code and classify these data
- There are broad gains for patients once information is extracted from these data
- The clinicians who generate these notes have first-rate input to how to automatically mine these data using AI

In the next few slides I'll show you some examples from actual projects that we have done and also try to explain how we have done them

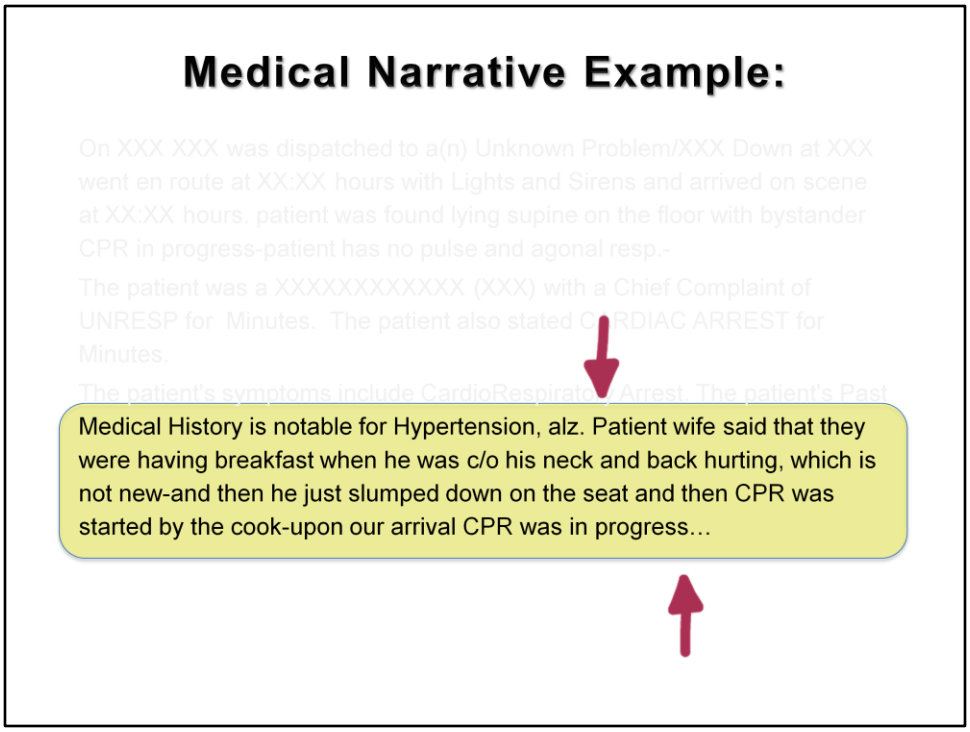
Medical Narrative Example:

On XXX XXX was dispatched to a(n) Unknown Problem/XXX Down at XXX went en route at XX:XX hours with Lights and Sirens and arrived on scene at XX:XX hours. patient was found lying supine on the floor with bystander CPR in progress-patient has no pulse and agonal resp.-

The patient was a XXXXXXXXXXXX (XXX) with a Chief Complaint of UNRESP for Minutes. The patient also stated CARDIAC ARREST for Minutes.

The patient's symptoms include CardioRespiratory Arrest. The patient's Past Medical History is notable for Hypertension, alz. Patient wife said that they were having breakfast when he was c/o his neck and back hurting which is not new-and then he just slumped down on the seat and then cpr was started by the cook-upon our arrival cpr was in progress…

This is a snippet from an ambolance report from a project we did in New Mexico. The goal was to optimize the allocation of equipment for cardiac arrest treatment on Ambulances and in public spaces.
Pay attension – it is really a story**<click>**



<read snippet from slide>**<click>**

Important parameters in Cardiac Arrest event:

Cardiac Arrest Data Collection Form

Date of arrest YYYY/MM/DD

Patient identifier (first name, last name, or ID number)

Sex

Age years (estimated) **OR** Date of birth YYYY/MM/DD

Cardiac arrest determined by

Cause of arrest

Treatment before EMS arrival

Bystander CPR

Defibrillation by bystander ☐ or implanted defibrillator ☐

Resuscitation attempted by EMS

Location of arrest out of hospital in hospital

Witnessed If witnessed, time of arrest HH:MM

Initial rhythm

Witnesses? (YES / NO)

CPR by Bystander? (YES / NO)

AED used?

We defined a set of important parameters to extract for each Ambulance Report. For example: <read first 3>**<click>**

CPR = Cardiopulmonary resuscitation

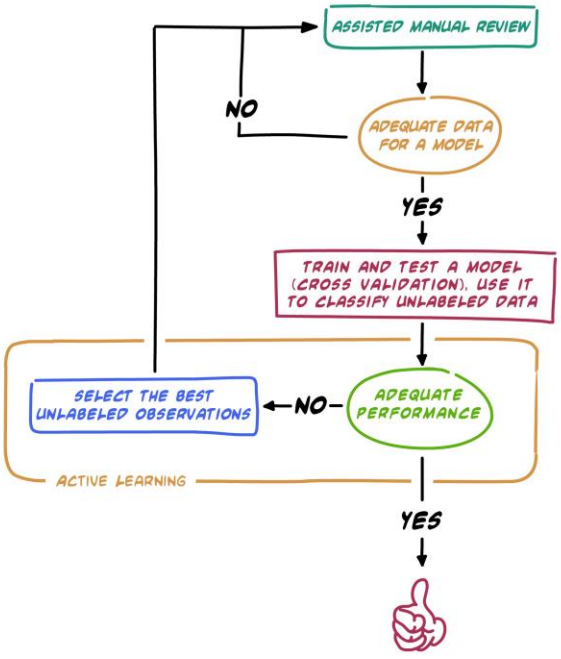
automated external defibrillator (**AED**)

Using our AI intelligence workflow, we automatically extracted those parameters from thousands of ambulance reports.

How have we done it? Let me explain our common workflow:

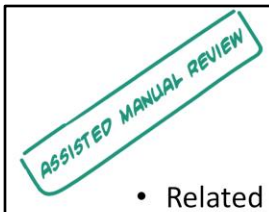
AI
Workflow

AI
Workflow


ASSISTED MANUAL REVIEW
NO
ADEQUATE DATA FOR A MODEL
YES
TRAIN AND TEST A MODEL (CROSS VALIDATION), USE IT TO CLASSIFY UNLABELED DATA
SELECT THE BEST UNLABELED OBSERVATIONS
NO
ADEQUATE PERFORMANCE
ACTIVE LEARNING
YES

Text Highlights

History of Present Illness
XXXXXXX with a past medical history significant for 1q43 chromosomal deletion, cerebral palsy seizure disorder, and aspiration pneumonia complicated by Pseudomonal infection who presented with fever, cough, respiratory distress, and persistent desaturations. Recent diagnosis penumonia 9 days ago, started on ciprofloxacin. Initiallly seemed improved but developed increased WOB, junkiness, then fever 103.2 Today, had desats 80s (not usually on home 02). Brought to the ED for evaluation. Of note, he had been tolerating his G-tube feeds without issue and has maintained normal urine output. No known sick contacts. No diarrhea, and new rashes.



- Related ICD 9 / 10
- Spelling mistakes / conventions:

Ex	Word	Count	% Count	Docs	% Matching Docs	% All Docs
☐	Seizure	382	15,77	158	23,72	9,44
☐	seizure	1256	51,86	237	35,59	14,17
☐	seizures	603	24,90	169	25,38	10,10
☐	SZ	6	0,25	4	0,60	0,24
☐	sz	88	3,63	40	6,01	2,39
☐	Seizures	23	0,95	19	2,85	1,14
☑	seized	5	0,21	4	0,60	0,24
☐	seizing	35	1,45	18	2,70	1,08
☑	seize	11	0,45	7	1,05	0,42
☐	seizuer	3	0,12	2	0,30	0,12
☐	seizes	2	0,08	2	0,30	0,12
☐	Seizing	1	0,04	1	0,15	0,06
☐	seizues	1	0,04	1	0,15	0,06
☑	Seized	1	0,04	1	0,15	0,06

- Related conditions: e.g. Postictal state
- Related examination: e.g. EEG (Electroencephalography)
- Related drugs: e.g. Phenobarbital
- Relevant lookarounds:
  "Seizure in the newborn“
  “Prolonged seizure”
- Negation forms:
  "No seizure"
  "No ongoing seizure"
- Ranking based on aggregated semantic, spelling and phonology rules

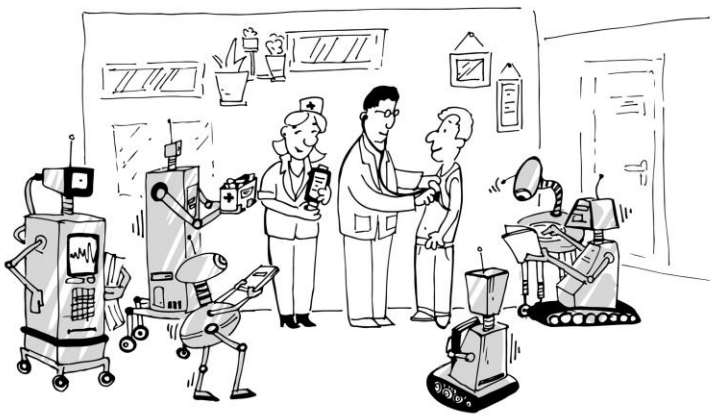
The optimization objective of our AI workflow is the time it takes a clinician to answer a clinical question using big data

Example: Peripheral IV infiltrations

- Infiltrations are patient safety events that need to be reported. Unfortunately they are under reported.
- We want to find out the ratio of unreported events
- The initiative and research is done by a nurse

- Preliminary dataset of 21k inpatient notes
- 2 workflow iterations
- 23% yied from high ranked manually reviwed notes
- 96 infiltrate events id'ed; 43 not reported
- 15 hours over 10 days (including SW installation and user training)

Realizing our vision



Currently, the big progress and breakthrough in AI and NLP is not mirrored by similar breakthroughs in clinical care.

It is only by making these methods available for direct interaction with clincians and researchers,
and by harnessing their expertise will the world of medicine benefit from AI.

We are just at the beginning…

While our work with clinicians is in its early stages, the impact on medical care has been tremendous:

- Febrile seizures
- Reducing unnecessary CT scans in children
- Changes in industrial standards
- List of publications: www.documentreviewtools.com

1. We are no longer doing lumbar punctures or CTs on children with febrile seizures
2. We are no longer using IV contrast CT scan on children with penetrating oral trauma
3. we have identified dangerous products that may injure children, leading to changes in industry standards