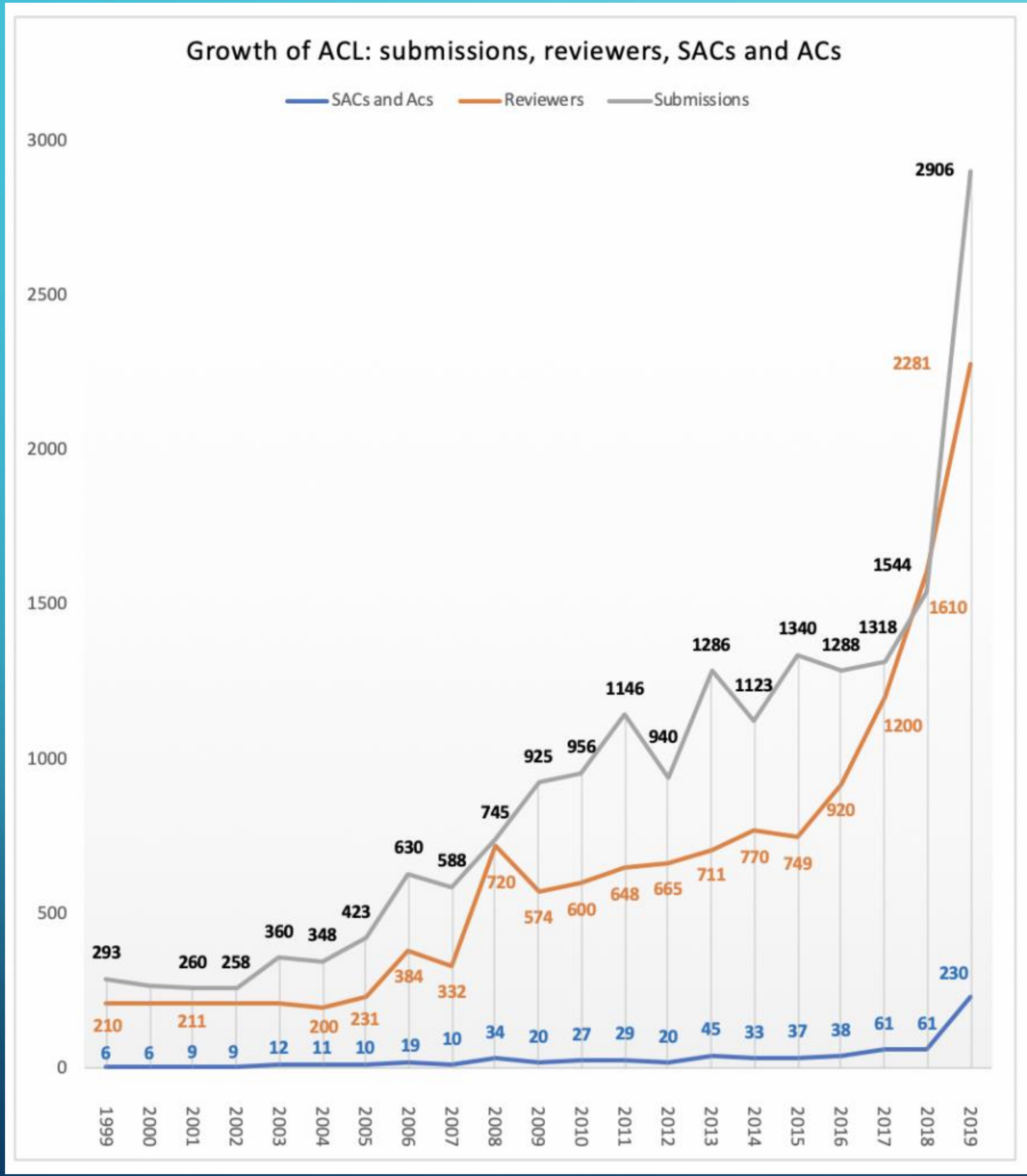
NN IS GREAT, BUT NLP IS NOT EQUAL TO NN

FEI XIA

UNIVERSITY OF WASHINGTON

NLPCC, OCT 14, 2019



ACL 2019 Submission:

- A 75% increase over ACL 2018.
- 10 times over ACL 1999.

A large majority of the papers use DNN

NLP IN OUR DAILY LIFE

- Email filters
- Predictive text: autocorrect
- Search engine: Baidu, Google, Bing
- Smart assistants and chatbots: Siri, Alexa
- Targeted ads
- Data analysis
- MT systems

CHALLENGES

- Interpretability
- Amount of data required
- Computationally expensive
- Need to tune hyperparameters

Much discussion on those challenges:

- Dr Ming Zhou's ACL2019 presidential address
- Dr. Keh-Yih Su's NLPCC2019 keynote speech

- Commonsense reasoning
- Combine knowledge and DNN
- Discourse and dialogues
- …

A NEW PARADIGM FOR NLP RESEARCH

- Pick a task
- Browse latest papers to find ones with the best results
- Download the source code and data
- Try different DNN architectures/pre-trained models on the dataset
- Modify the models and tune hyperparameters (until beating the SOTA or cutting the loss)
- Write a paper and submit

THE NEW PARADIGM (CONT)

- Quick development cycles, leading to more publications
- SOTA are constantly improved.
- Easy to enter the NLP field.
- Treating NLP tasks as pure ML problems:
  - Assuming the task is well defined: i.e., it is clear what the input x and the output y are. we just need to find a function that maps x to y.
  - Assuming training/test data are available (and are reasonably large).
  - For ML, use DNN almost exclusively.
- It is often not clear where the improvement comes from: pre-trained models, DNN architecture, hyperparameter tuning, idiosyncrasy of the dataset.

NLP $\neq$ NN

- NLP $\neq$ ML:
  - NLP is about understanding, processing or generation of natural language text.
  - When applying NLP to a particular domain, we often need to understand that domain.
  - As NLP techniques have been applied to our daily activities, their social impact should be considered.

THE WINOGRAD SCHEMA CHALLENGE (2016)

- The trophy doesn't fit into the brown suitcase because **it** was too large/small.
- Joan made sure to thank Susan for all the help **she** had received/given.
- Paul tried to call George on the phone, but **he** wasn't successful/available.
- The lawyer asked the witness a question, but **he** was reluctant to repeat/answer it.

NLP ≠ NN

- NLP ≠ ML:
  - NLP is about understanding and processing natural languages, and hopefully also about languages themselves.
  - When applying NLP to a highly professional domain, we often need to understand that domain.
  - As NLP techniques have been applied to our daily activities, their social impact should be considered.
  - …

- ML ≠ NN:
  - There is no free lunch.
  - Like AI, NN has gone through several boom-bust cycles.

AI IN THE MEDIA

- “Robots will destroy our jobs – and we’re not ready for it” (The Guardian, 1/11/2017)
- “Robots may replace 800 million workers by 2030” (CNBC, 11/30/2017)
- “How frightened should we be of AI?” (The New Yorker, 5/14/2018)
- “AI Transforming the World” (Forbes, 2/24/2019)
- “Learning to love the AI Bubble” (MIT Sloan Review, 5/15/2019)
- “Big Data is Dead. Long Live Big Data AI” (Forbes, 7/1/2019)
- “The AI tech bubble will burst if it can’t prove itself beyond hype” (Health Europa, 8/27/2019)
- “AI: It’s complicated and unsettling, but inevitable” (Forbes, 9/10/2019)

HISTORY OF AI

- 1952-1956: The birth of AI

- 1956-1974: The golden years
- 1974-1980: The 1st AI winter

- 1980-1987: Boom
- 1987-1993: The 2nd AI winter

- 1993-2011: Began to be used successfully in the IT industry
- 2011-now: Deep learning, big data, and artificial general intelligence

NLP ≠ NN

- NLP ≠ ML:
  - NLP is about understanding and processing natural languages, and hopefully also about languages themselves.
  - When applying NLP to a highly professional domain, we often need to understand that domain.
  - As NLP techniques have been applied to our daily activities, their social impact should be considered.
  - …

- ML ≠ NN:
  - There is no free lunch.
  - Like AI, NN has gone through several boom-bust cycles.

OUTLINE

- Understanding the domain: Clinical NLP as a case study
- Social impact of NLP
- Understanding languages: NLP and linguistics (if time permits)

UNDERSTANDING THE DOMAIN: CLINICAL NLP AS A CASE STUDY

BIONLP

- Clinical NLP:
  - Focus on documentation related to patient care
  - Electronic Medical Records (EMRs)

- Biomedical NLP:
  - Focus on scientific discoveries about biology, physiology, and medicine
  - Journal articles, clinical trials, webpages, …

HEALTHCARE IN USA

- Healthcare spending is about 18% of GDP in 2018:
  - However, healthcare quality is at the bottom compared with other developed countries.
  - Many factors: for-profit nature of the health insurance industry, lawsuits, political fights.

- Clinical data:
  - Number of visits: 1.2 billion each year
  - Tons of EMR notes: more than 80% of data are free text.

- Clinical NLP has a huge potential.

ELECTRONIC MEDICAL RECORDS (EMRS)

- Most healthcare facilities in the US store medical records in electronic forms, using medical systems such as Epic.
  - Percent of office-based physicians using any EMR system: 78.4% (2013), 86% (2017).

- The medical records contain:
  - Structured data: problem lists, lab results, pharmacy orders, etc.
  - Unstructured data (Free-form text): radiology reports, operative notes, discharge summaries, …
  - More than 80% of data is unstructured one.

- There are different kinds of medical records: e.g., radiology reports, admit notes, operative notes, discharge summaries, etc.

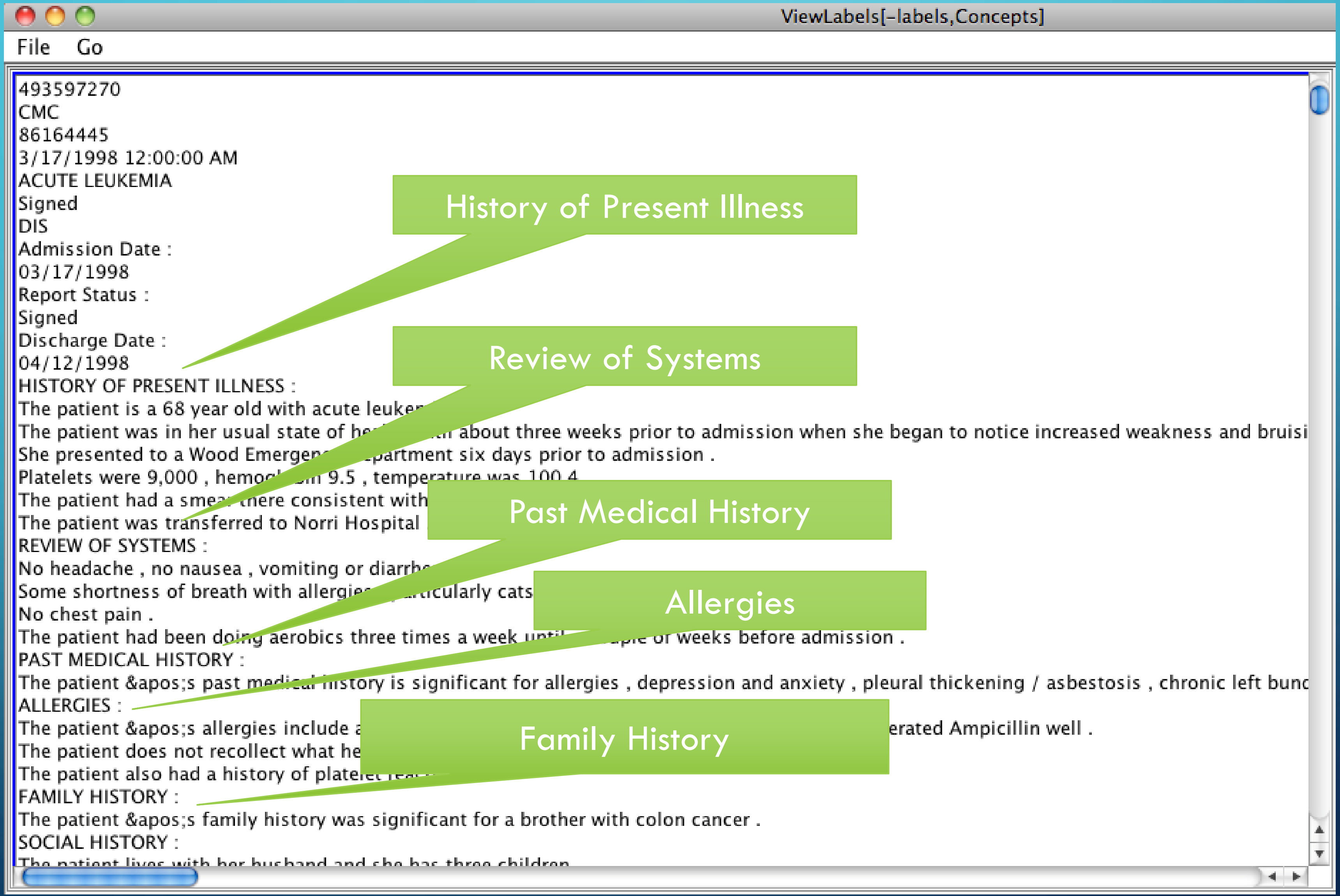
A DISCHARGE SUMMARY

ViewLabels[-labels,Concepts]

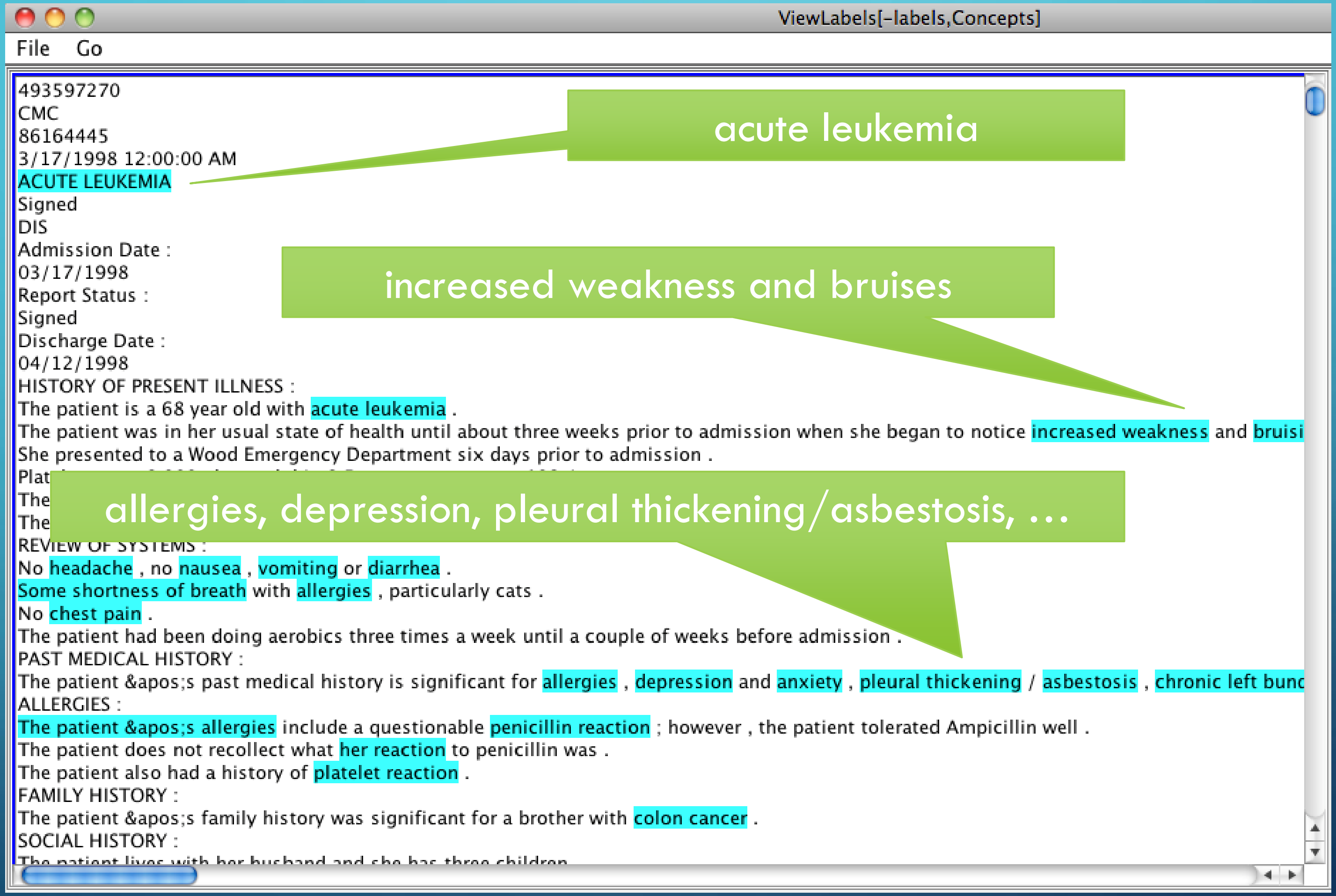
File Go

493597270
CMC
86164445
3/17/1998 12:00:00 AM
ACUTE LEUKEMIA
Signed
DIS
Admission Date :
03/17/1998
Report Status :
Signed
Discharge Date :
04/12/1998
HISTORY OF PRESENT ILLNESS :
The patient is a 68 year old with acute leukemia .
The patient was in her usual state of health until about three weeks prior to admission when she began to notice increased weakness and bruisi
She presented to a Wood Emergency Department six days prior to admission .
Platelets were 9,000 , hemoglobin 9.5 , temperature was 100.4 .
The patient had a smear there consistent with ALL .
The patient was transferred to Norri Hospital .
REVIEW OF SYSTEMS :
No headache , no nausea , vomiting or diarrhea .
Some shortness of breath with allergies , particularly cats .
No chest pain .
The patient had been doing aerobics three times a week until a couple of weeks before admission .
PAST MEDICAL HISTORY :
The patient 's past medical history is significant for allergies , depression and anxiety , pleural thickening / asbestosis , chronic left bunc
ALLERGIES :
The patient 's allergies include a questionable penicillin reaction ; however , the patient tolerated Ampicillin well .
The patient does not recollect what her reaction to penicillin was .
The patient also had a history of platelet reaction .
FAMILY HISTORY :
The patient 's family history was significant for a brother with colon cancer .
SOCIAL HISTORY :
The patient lives with her husband and she has three children

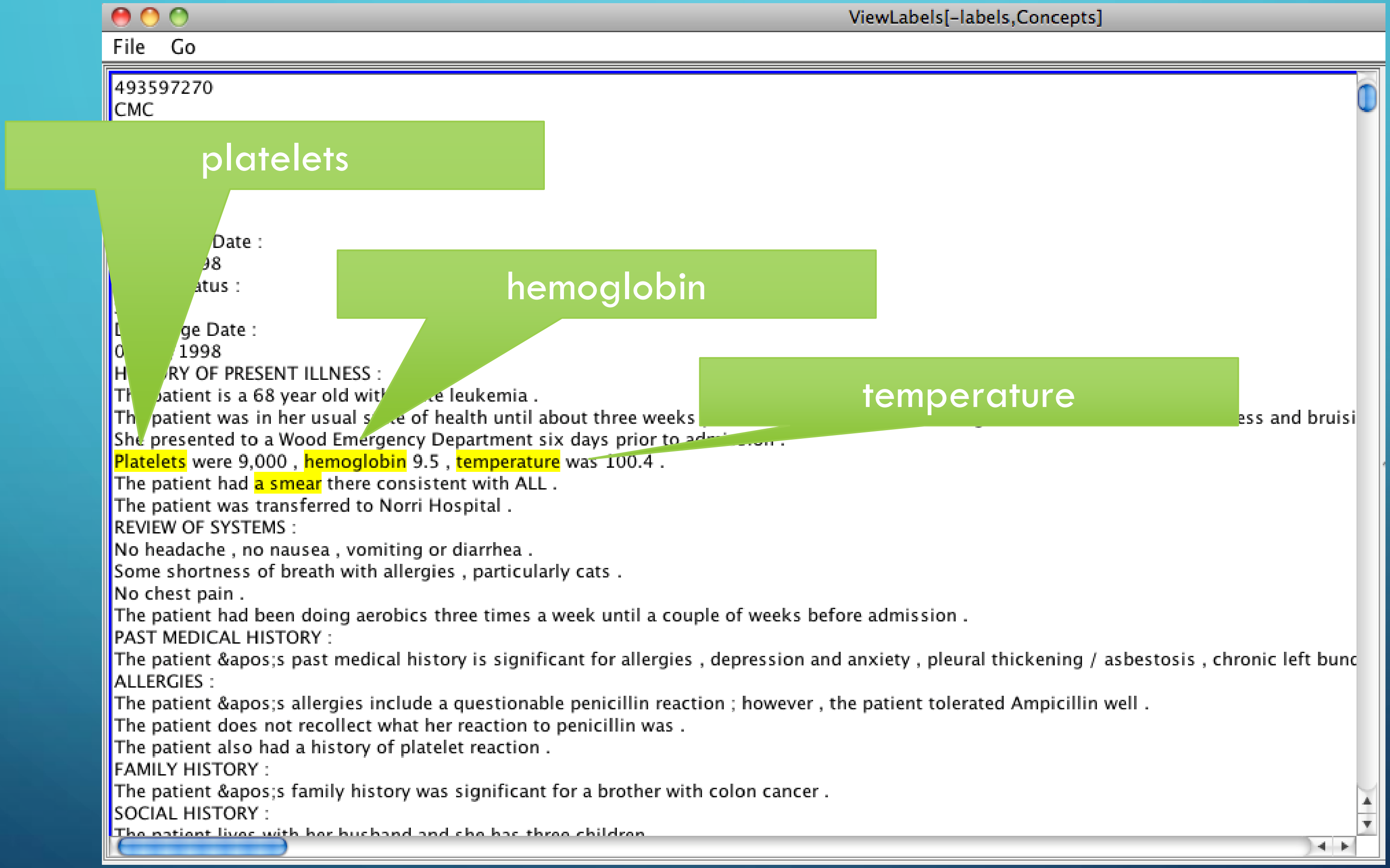
SECTIONS


ViewLabels[-labels,Concepts]
File Go
493597270
CMC
86164445
3/17/1998 12:00:00 AM
ACUTE LEUKEMIA
Signed
DIS
Admission Date :
03/17/1998
Report Status :
Signed
Discharge Date :
04/12/1998
HISTORY OF PRESENT ILLNESS :
The patient is a 68 year old with acute leuke
The patient was in her usual state of he about three weeks prior to admission when she began to notice increased weakness and bruisi
She presented to a Wood Emergen partment six days prior to admission .
Platelets were 9,000 , hemog 9.5 , temperature was 100.4
The patient had a smea consistent with
The patient was transferred to Norri Hospital
REVIEW OF SYSTEMS :
No headache , no nausea , vomiting or diarrh
Some shortness of breath with allergie cularly cats
No chest pain .
The patient had been doing aerobics three times a week until of weeks before admission .
PAST MEDICAL HISTORY :
The patient 's past medi history is significant for allergies , depression and anxiety , pleural thickening / asbestosis , chronic left bunc
ALLERGIES :
The patient 's allergies include erated Ampicillin well .
The patient does not recollect what he
The patient also had a history of plate
FAMILY HISTORY :
The patient 's family history was significant for a brother with colon cancer .
SOCIAL HISTORY :
History of Present Illness
Review of Systems
Past Medical History
Allergies
Family History

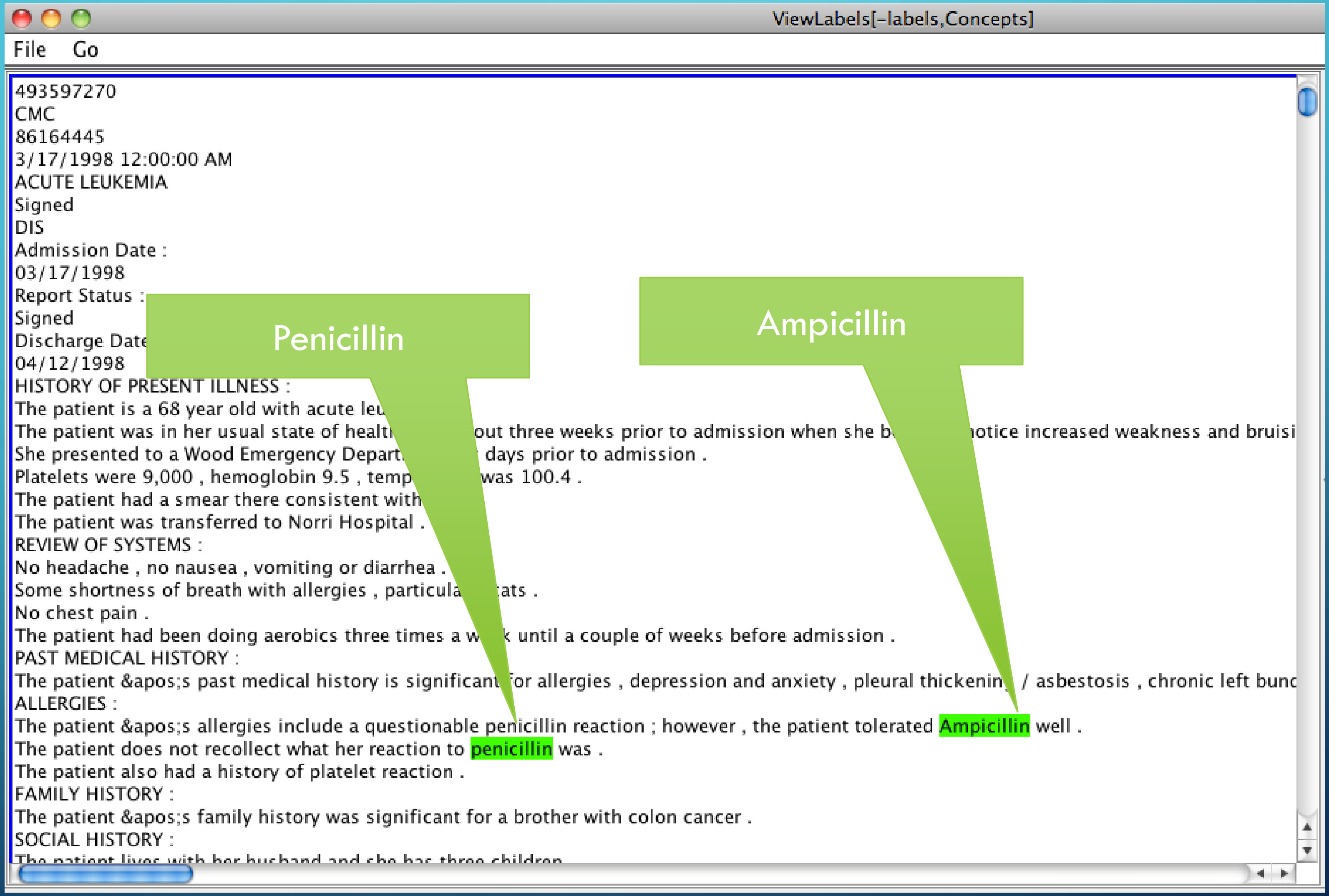
MEDICAL PROBLEMS



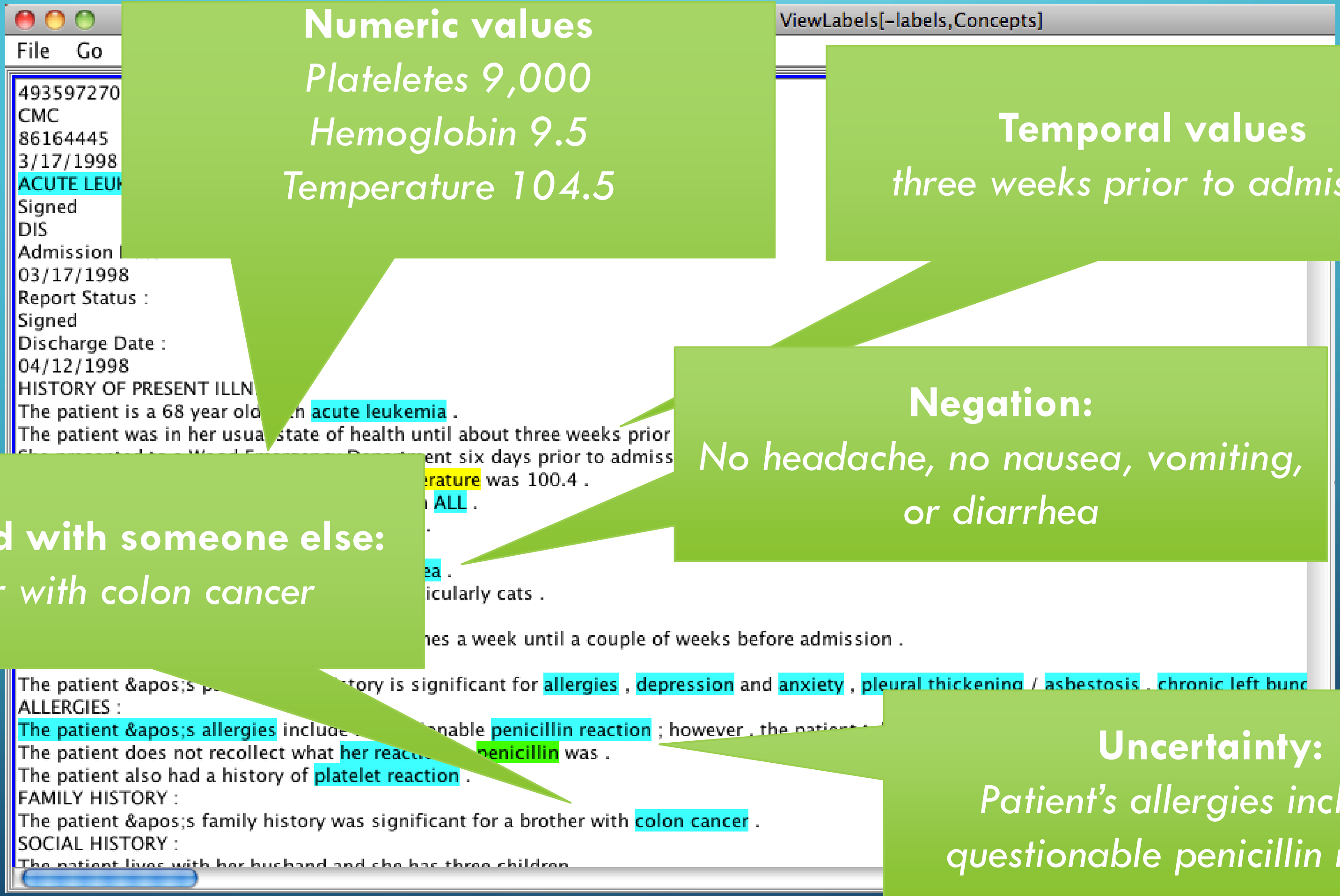
MEDICAL TESTS



MEDICAL TREATMENTS



OTHER TYPES OF INFORMATION



Numeric values

Plateletes 9,000

Hemoglobin 9.5

Temperature 104.5

Temporal values

three weeks prior to admission

Negation:

No headache, no nausea, vomiting, or diarrhea

Associated with someone else:

brother with colon cancer

Uncertainty:

Patient's allergies include a questionable penicillin reaction

NLP APPLICATIONS IN THE CLINICAL DOMAIN

- Cohort selection, phenotyping:
  - Finding groups of patients that satisfy particular criteria
  - Does/Did the patient have X?
- Decision support systems:
  - What stage of cancer is this patient in?
  - What are common practices for treating this disease? What is the success rate for each practice? What is the best practice under this condition (i.e., precision medicine)?
- Quality of patient care:
  - What is the rate of catching hospital-acquired pneumonia (HAP) in this hospital?
  - Is this patient likely to develop HAP in the next 48 hours?
- …

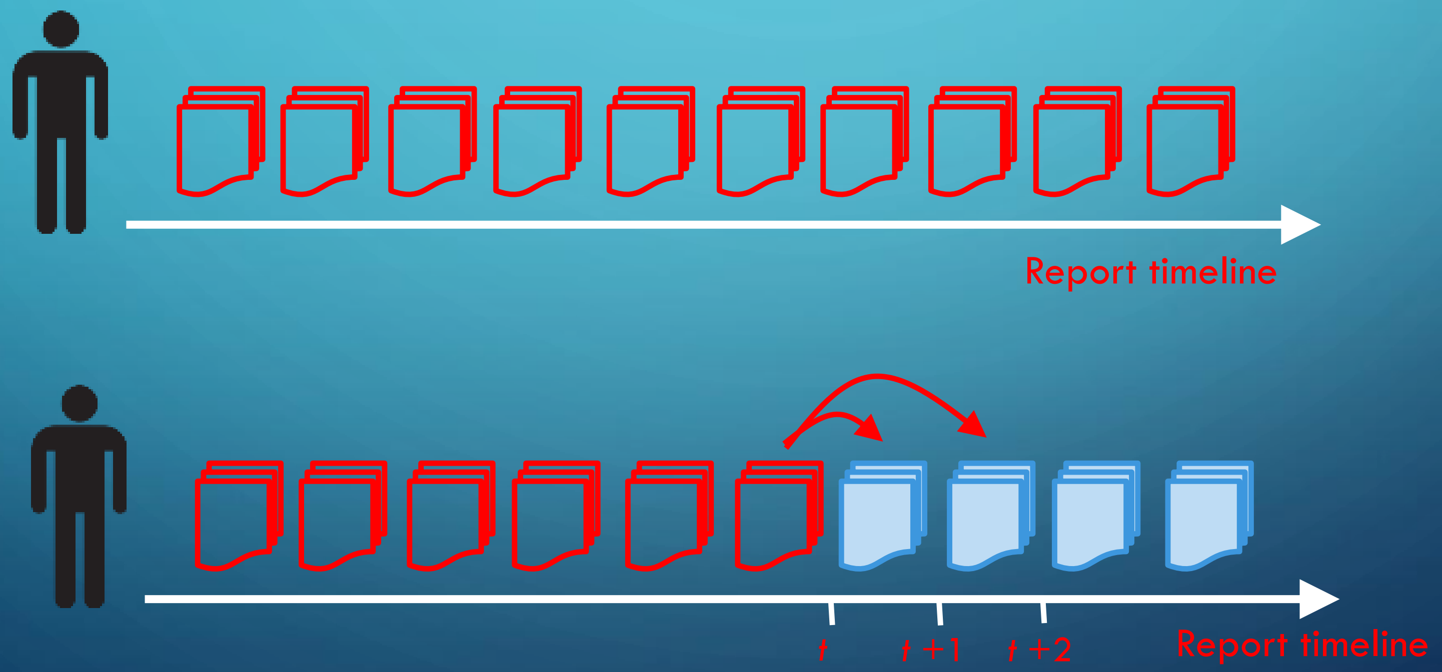
SOME OF OUR RECENT PROJECTS

(1) Radiology recommendation extraction:

- To find clinically important recommendation made by the radiologist in a given radiology report to advise the referring clinician to further evaluate an imaging finding
- Failure to communicate of abnormal test results is the #2 cause of radiology malpractice lawsuits.
- Many errors are due to hedging (an evasive statement to avoid the risk of commitment)
- Ex: "*If clinically indicated, pelvic ultrasound could be performed in 4 to 6 weeks to document resolution.* "

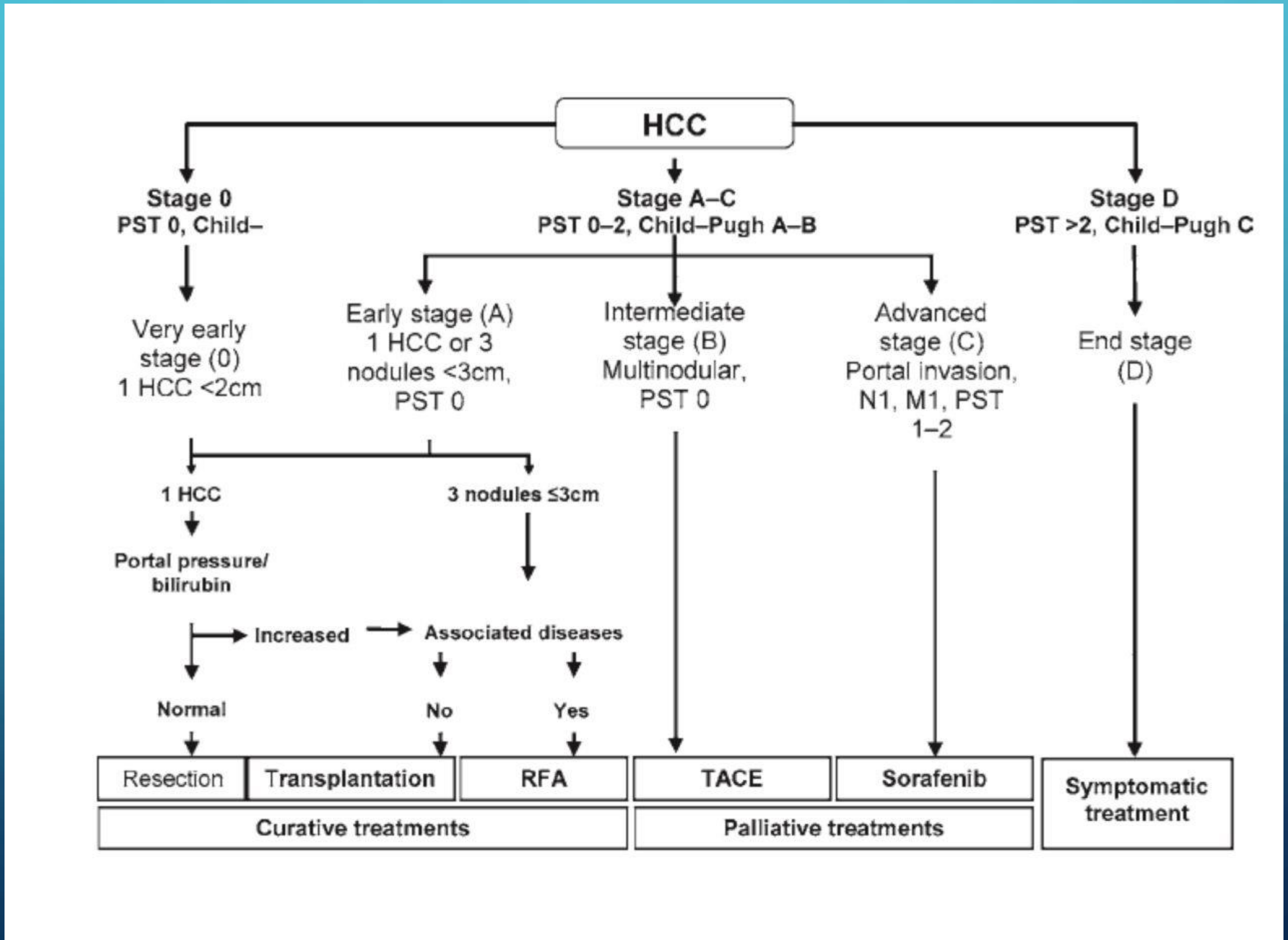
(2) Critical illness phenotype detection and prediction:

- A phenotype is any observable characteristic or trait of a disease
- Ex: Hospital-acquired pneumonia, acute lung injury
- Three settings: had/has/will-have phenotype



(3) Hepatocelluar carcinoma (HCC) staging:

- HCC stages are based on numbers and sizes of tumors, regional lymph nodes, distant metastasis, etc.



CHALLENGES IN CLINICAL NLP

- Formulate the problem:
  - Ex: Phenotype detection and HCC staging: a classification task or an IE task followed by heuristics?
- Data:
  - Getting access to the clinical data: privacy issue
  - Creating gold standard: "Expert annotation"
  - Imbalanced data sets
- Output:
  - Interpretability of system output
  - Proper dissemination of results: understand the complication and consult medical experts

CHALLENGES IN CLINICAL NLP (CONT)

- Requiring "deep" understanding:
  - Ex: "No PNA detected", "PNA is likely", "two brothers, … one had PNA"
  - Ex: "The chest is otherwise unchanged"
- Hedging: how to determine whether some statement is hedging?
- Incorporating domain knowledge: e.g., ULMS
- Scalability
- The role of medical experts

CHALLENGES IN CLINICAL NLP (CONT)

- Requiring “deep” understanding:
  - Ex: “No PNA detected”, “PNA is likely”, “two brothers, … one had PNA”
  - Ex: “The chest is otherwise unchanged”
- Hedging: how to determine whether some statement is hedging?
- Incorporating domain knowledge: e.g., ULMS
- Scalability
- The role of medical experts

MEDICAL EXPERTS

- Identify a problem that NLP can potentially be useful
- Choose the raw data for the task
- Annotate the data (if needed)
- Help NLP researchers to determine relevant domain knowledge
- Interpret system results
- Choose appropriate ways to disseminate the results

MEDICAL EXPERTS

- Identify a problem that NLP can potentially be useful
- Choose the raw data for the task
- Annotate the data (if needed)
- Help NLP researchers to determine relevant domain knowledge
- Interpret system results
- Choose appropriate ways to disseminate the results

A FEW CORPORA IN OUR BIONLP PROJECTS

Corpus	Report Type	Corpus Size	Annotation	Annotation Unit	Annotators
C1	Radiology reports	800 reports	Critical recommendation	Sentence	1 radiologist 1 internal medicine physician
C2	Chest x-ray reports	1344 reports	PNA and CPIS	Report	1 surgeon 1 data analyst
C3	Eight ICU report types	5313 reports for 426 patients	PNA	Patient	1 research RN

HOW OFTEN DO EXPERTS AGREE?

- Clinical Pulmonary Infection Score (CPIS):
  - 1A: no infiltrate
  - 1B: Diffuse infiltrate or atelectasis
  - 1C: Localized infiltrate
- Two annotators:
  - A1: A surgeon
  - A2: A data analyst
- Annotation:
  - 100 reports
  - Two rounds: without and with annotation guidelines

INTER-ANNOTATOR AGREEMENT – CHEST X-RAY CORPUS

Clinical Pulmonary Infection Score (CPIS)

Round	A1	A2	Agreed	Acc	Kappa
1	13/59/28	15/74/11	12/52/6	70%	0.415
2	13/72/15	16/72/12	13/68/10	91%	0.797

Pneumonia (PNA)

Round	A1	A2	Agreed	Acc	Kappa
1	13/59/28	15/74/11	12/52/6	45%	0.085
2	67/19/15	67/32/1	13/68/10	85%	0.697

IAA on **CPIS and PNA** labeling for the 100 double annotated reports in the chest X-ray corpus (C2).

- x/y/z in each cell of the "A1", "A2", and "Agreed" columns are the numbers of reports with labels 1a, 1b, and 1c, respectively

ANNOTATION GUIDELINE FOR CPIS (ROUND 2)

1A: No Infiltrate
• The report includes information that neither diffuse nor localized infiltrate. The report could include edema or pleural effusion. • If there are extra pleural mentions in the report, they are not related to PNA.
1B: Diffuse Infiltrate or Atelectasis
• Atelectasis is more important than localized process that is consistent with infection. • Lobar collapse is consistent with atelectasis. • Multiple areas of opacity could fall under 1B. • If bi-basilar consolidation is present with bi-pleural effusion much more suggestive of atelectasis.
1C: Localized Infiltrate
• If one opacity is specifically highlighted and PNA or infection also mentioned in text, than this is more important than 1A and 1B.

Medical conversation

Video-recordings

Transcripts

Annotations

PHASE: DIAGNOSIS

SEQ 1

PHASE: TREATMENT

SEQ 1

SEQ 2

TID	PR	TEXT
1	D	感冒 了. (He's) got a cold.
2	M	嗯. Ok.
3	D	吃点药吧? Take some oral medicine, ok?
4	M	啊? Huh?
5	D	先吃点药，好吧? Take some oral medicine first, ok?
6	M	哦，好. Oh, ok.
7	D	嗯. Ok.
8	M	开 点 青霉素 给 我们 好吧? Could you prescribe us some Penicillin?
9	D	嗯，行. Ok. Alright.

Item	Number
Total Number of Visits	318
Total Number of Hospitals	5
Total Number of Physicians	9
Total Number of Patients	318
Average length of a visit	4.9 minutes
Total length of the recordings	26 hours

Item	Total	Avg.
Characters	468,162	1472.2
Words	270,042	849.2
Turns	39,216	123.3
Non-verbal turns	5,815	18.3
Adjacency pairs	20,123	63.3
Sequences	9114	28.7

Main outcome (Wang et al, 2018)

- COSTA: a new scheme for conversation
- An annotated corpus
- A sociological study about caregiver actions and medicine prescribing decisions

PUBLIC HEALTH PLATFORM

请在此输入您的问题
搜索
向医生提问
搜药
登录 / 注册
胃肠内科 血液科专家 医生免费咨询 胃肠怎么查 精神科医生 消化内科哪里好 中医软件 神经内科的专家
免费 的服务器 神经内科专家 神经内科排名 胃肠专家 在线心理医生 天坛神经内科 秃头挂什么科 精神科 快速入睡方法
老龄斑怎么去掉 怎么去掉老人斑 什么症状 美胸整形 内科宣传栏 女人光子嫩肤 39wenyis 内科湿疹 脑病有哪些
首页 > 内科
肾内科 消化内科 内分泌科 血液科 风湿免疫科 呼吸内科 神经内科 心血管内科
最新问题
我要提问
一到晚上就喘不上来大气，就憋着很难受
朱立萍 助理医师
是否是夜间睡觉突然憋醒，然后坐起来感觉好一点，如果是这样的话，建议去医院看心内科门诊，做个心电图，拍个胸片排除心肺方面的疾病。如果不是上述情况的话，是否和房间门窗关闭不透风有关，是否和心理压力有关。是否是夜间睡觉突然憋醒，然后坐起来感觉好一点，如果是这样的话，建议去医院看心内科门诊，做个心电图，拍个胸片排除心肺方面的... [详细]
感觉自己的腿有点虚，走路姿势也跟别人不
朱立萍 助理医师
可能是平常锻炼的少了，要经常锻炼身体，如果还是不行就建议去医院检查一下肌电图，看看是不是肌无力。可能是平常锻炼的少了，要经常锻炼身体，如果还是不行就建议去医院检查一下肌电图，看看是不是肌无力。可能是平常锻炼的少了，要经常锻炼身体，如果还是不行就建议去医院检查一下肌电图，看看是不是肌无力。
推荐医生
更多
吴化忠 主治医师 提问
安徽合肥肥东县人民医院
杨学水 主治医师 提问
肥城陶阳矿医院
孔伟伟 医师 提问
晋城华肤皮肤病专科医院
王晋军 主治医师 提问
柳林县柳林镇卫生院
李艳芳 主管护师 提问
辉县市人民医院

THE CHIMED CORPUS (TIAN ET AL., 2019)

Department	内科 > 淋巴增生 Internal Medicine > Lymphocytosis
Title	胃部淋巴增生会癌变吗? Will lymphatic hyperplasia in the stomach cause cancer?
Question	我最近检查出患有胃部淋巴增生的疾病，非常担心，请问它会癌变吗? I recently checked out the disease of lymphoid hyperplasia in the stomach. I am very worried. Will it cause cancer?
Keypharses	慢性浅表性胃炎，幽门螺旋杆菌感染，淋巴增生，胃，消化 Chronic superficial gastritis, Helicobacter pylori infection, lymphatic hyperplasia, stomach, digestion
Answer 1	这一般是幽门螺旋杆菌感染造成的，一般不会造成癌变，所以不必惊慌。建议饮食规律，吃易消化的食物，细嚼慢咽，少量多餐，禁食刺激性食物。 In general, this is caused by Helicobacter pylori infection and does not cause cancer. So do not panic. It is recommended to have a regular diet, eat digest friendly food and chew slowly. Do not eat much in one meal and no spicy food is allowed.
Adopted	True
Answer 2	这是普通的慢性胃粘膜炎症，与幽门螺旋杆菌感染有关。可用阿莫西林治疗。 This is a common chronic gastric mucosal inflammation and has a relationship with Helicobacter pylori infection. You can choose amoxicillin for treatment.
Adopted	False

A QUICK SUMMARY SO FAR

- Clinical NLP has many real-life applications, but there are many challenges.
- Two major issues are:
  - The need of working closely with domain experts
  - Getting access of raw and annotated data:
    - Allow de-identified data to be shared by researchers?
    - Organize more shared tasks (like i2b2) in the clinical domains

  ➔Require support from NLP researchers, medical experts, funding agencies, and the public.
- Meanwhile, take advantage of applications that are less dependent on data and experts (e.g., public health platform)

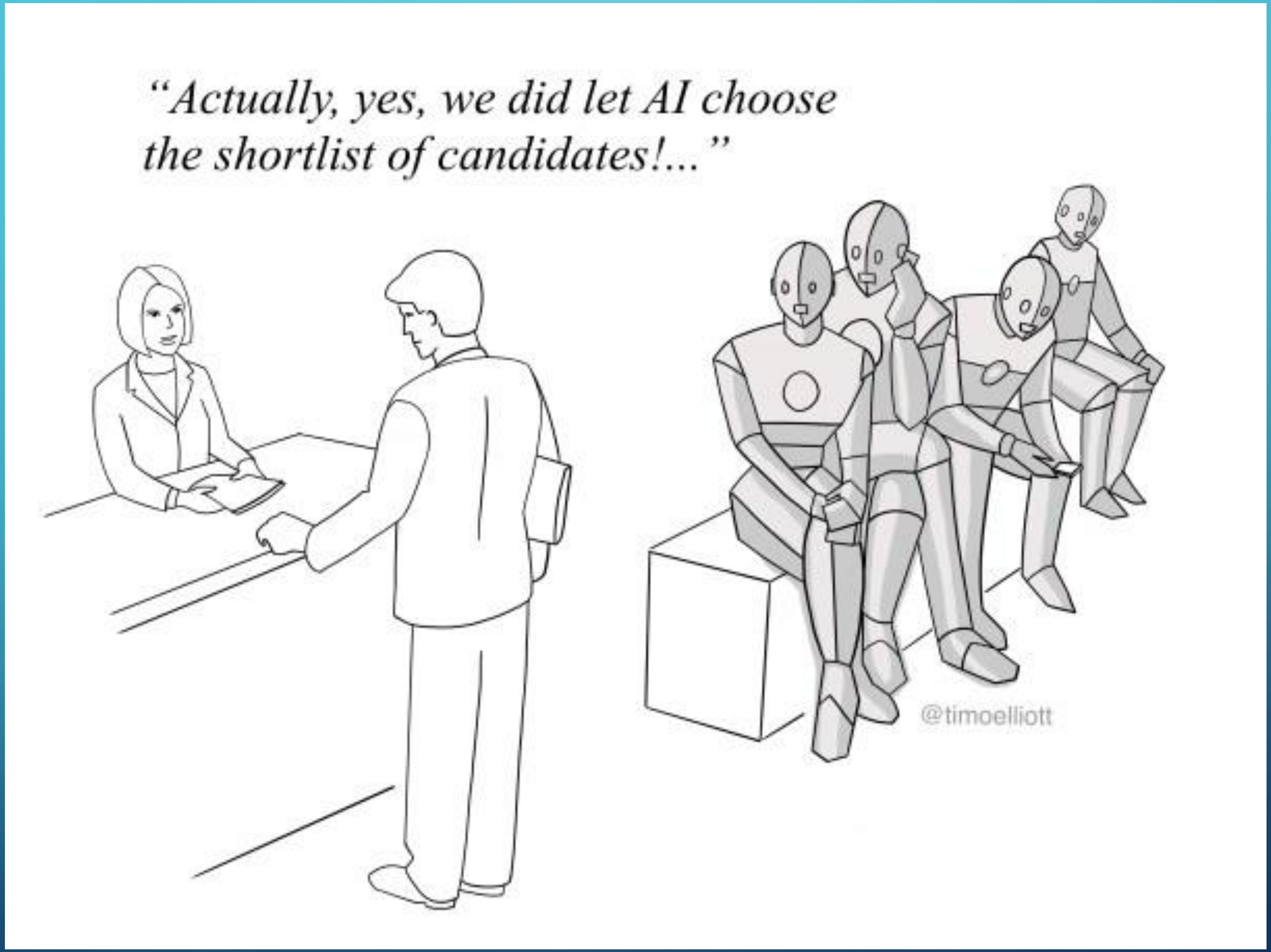
- Similar issues are likely to exist in other highly professional domains

SOCIAL IMPACT OF NLP

ETHICS IN NLP

- Ethics NLP workshops in 2017 and 2018.
- Ex: The social impact of NLP (e.g., Hovy and Spruit, 2016):
  - Exclusion: demographic bias of datasets
  - Overgeneralization of models
  - Overexposure and underexposure
- Ex: “Energy and policy considerations for DL in NLP” (Strubell et al., 2019)
- Ex: Gender/race bias in NLP systems
  - (Bolukbasi et al, 2016): “Man is to computer programmer as woman is to homemaker? Debiasing word embeddings”

Biased data lead to biased decisions



(This cartoon is from timoelliott.com)

IMAGE TO TEXT


©Bryant WWW.CARTOONaDAY.COM


imPEACH TRUMP

TOP 14 CAUSES OF POLITICAL POLARIZATION (BLANKENHORN, 2018)

- The rise of identify-group politics: left vs. right, liberal vs. conversative
- Geographical sorting: blue vs. red states
- Growing racial and ethnic diversity
- Growing religious diversity
- The spread of media ghettoes: the decline of the old media
- The growing influence of certain ways of thinking: e.g., confirmation bias
- …

CONFIRMATION BIAS

- The tendency to search for, interpret, favor, and recall information in a way that affirms one's prior beliefs or hypotheses:
  - Biased search for information
  - Biased interpretation
  - Biased memory

PERSONALIZED WEB SEARCH

- Web search that is tailored specifically to an individual's interests (using location, search history, past click-behavior, etc).

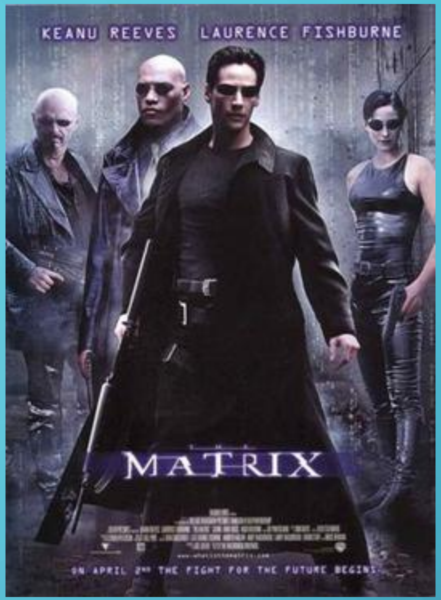
- Ex: Reranking the top N search results such that documents likely to be preferred by the user are presented higher.

- Preference: location (e.g., Chinese restaurant), occupation (e.g., Michael Collins), hobby (e.g., Eagles), ideology (e.g., progressive vs. conservative), etc.

TARGETED ADS

https://www.sticky.digital/danger-of-living-in-a-filter-bubble/


TARGETED ADS
PERSONAL DATA


KEANU REEVES LAURENCE FISHBURNE
MATRIX
ON APRIL 2ND THE FIGHT FOR THE FUTURE BEGINS

https://immediatefuture.co.uk/blog/what-should-we-think-of-filter-bubbles-on-social-media/

FILTER BUBBLE

- In Eli Pariser's 2011 book, "The Filter Bubble: What the Internet Is Hiding from You"

"A filter bubble is the intellectual isolation that can occur when websites make use of algorithms to selectively assume the information a user would want to see, and then give information to the user according to this assumption."

- *The problem becomes more severe on social media (e.g., Facebook).*

DANGERS OF FILTER BUBBLE

(Pariser, 2011):

- "The danger is that increasingly you end up not seeing what people who think differently see and in fact not even knowing that it exists"
- "Ultimately, democracy works only if we citizens are capable of thinking beyond our narrow self-interest. But…the filter bubble pushes us in the opposite direction — it creates the impression that our narrow self-interest is all that exists. And while this is great for getting people to shop online, it's not great for getting people to make better decisions together."

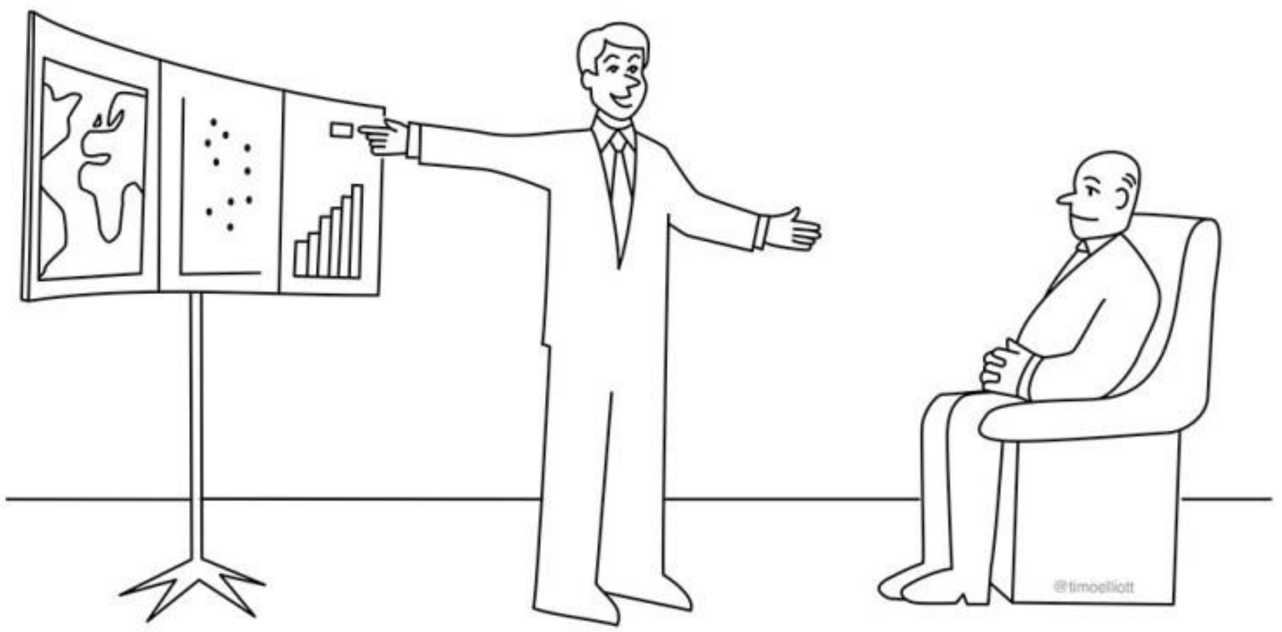
(Dylko et al, 2017; 2018):

- Study impact of customizability technology (e.g., personalized search) on political polarization
- Customizability technology has a strong effect on the minimization of exposure to counter-attitudinal information.

HOW TO “AVOID” FILTER BUBBLE

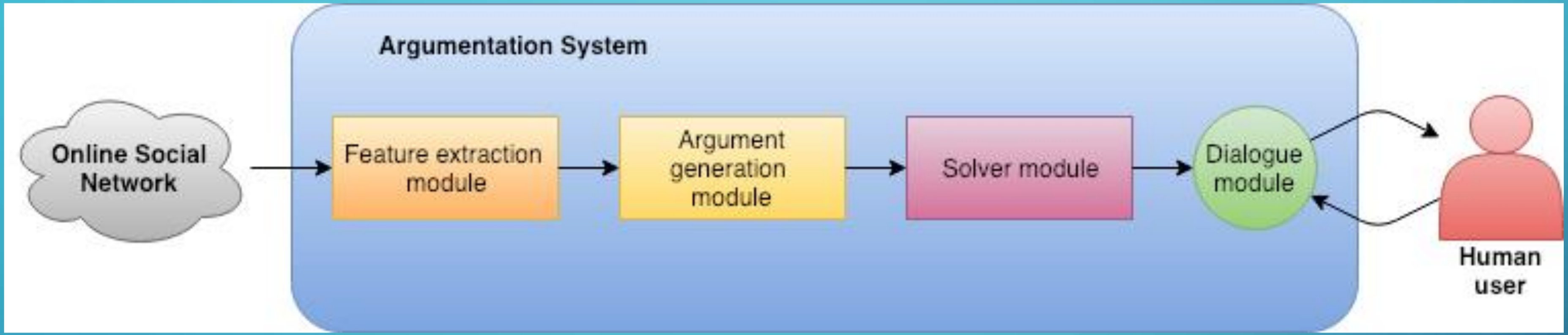
- Use incognito browsing, delete search history
- Delete or block browser cookies
- Use ad-blocking browser extensions
- Use the VPN
- Keep your social media data private and hide your birthday/gender

- Read news sites and blogs that provide a wide range of perspectives
- Talk and listen to people with different views
- Get touch with the outside world directly (e.g., travel, get offline periodically)



"And our unique JustifyIt™ feature uses deep learning to find data that agrees with your point of view!"

(This cartoon is from timoelliott.com)



Architecture of the argumentation system in (Ruiz-Dolz et al, 2019)

The real world

Exposure/exclusion

The internet

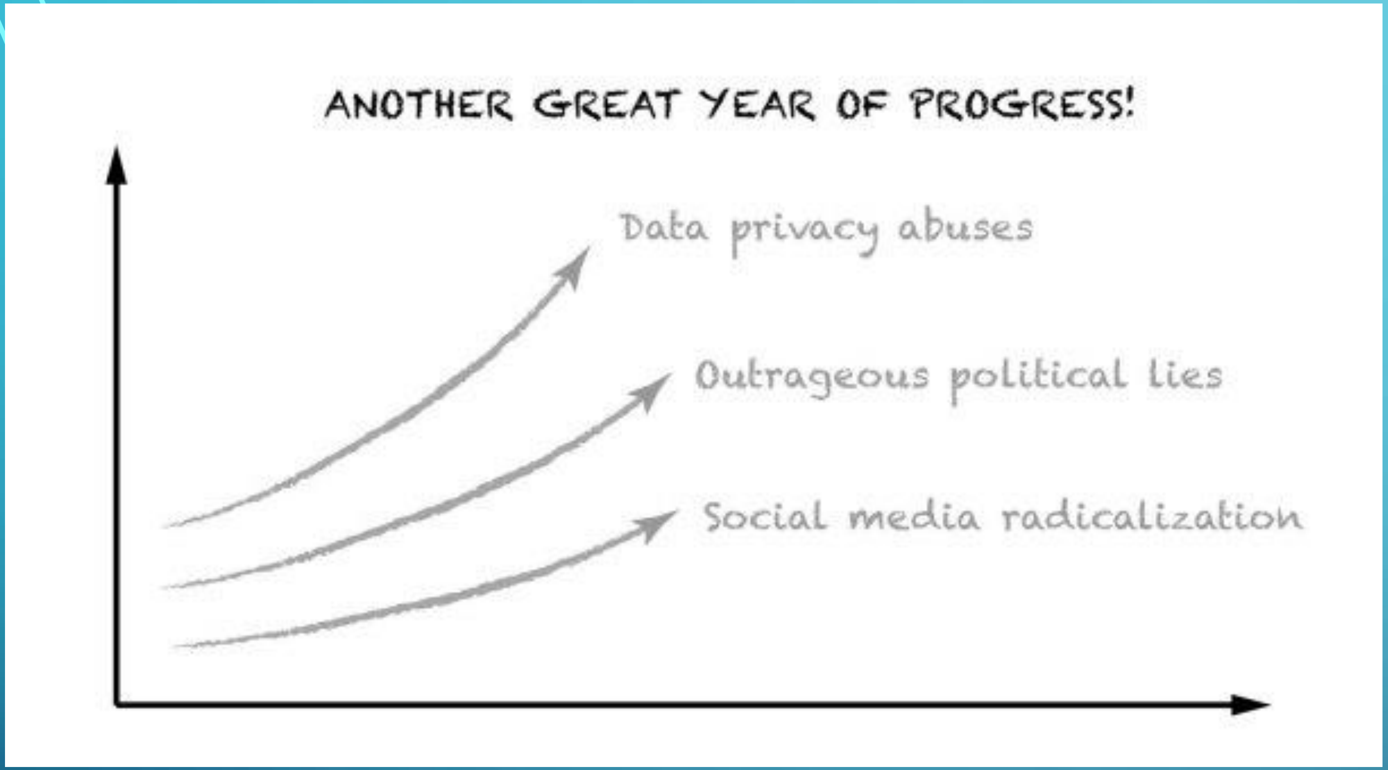
Personalized search and other NLP systems

Targeted content

Personal data

People

Social media



What's our role in this?

(This cartoon is from timoelliott.com)

CONCLUSION

- NLP has seen tremendous changes in the past few years:
  - SOTA improved constantly
  - NLP in IT industry, media, and daily life
  - New paradigm for NLP research
- However, NLP is not equal to ML in general (or NN in particular):
  - We need to understand the domain in order to incorporate domain knowledge: e.g., clinical NLP
  - Be aware of (potentially) negative social impacts of NLP.
  - We hope to understand more about human languages themselves: NLP and linguistics

"It was the best of times, it was the worst of times.

It was the age of wisdom, it was the age of foolishness.

It was the epoch of belief, it was the epoch of incredulity.

It was the season of Light, it was the season of Darkness.

It was the spring of hope, it was the winter of despair.

We had everything before us, we had nothing before us.

…."

(Charles Dickens, "A tale of two cities")

THANK YOU