

Topics in Machine Learning Machine Learning for Healthcare



Rahul G. Krishnan
Assistant Professor

Computer science & Laboratory Medicine and Pathobiology

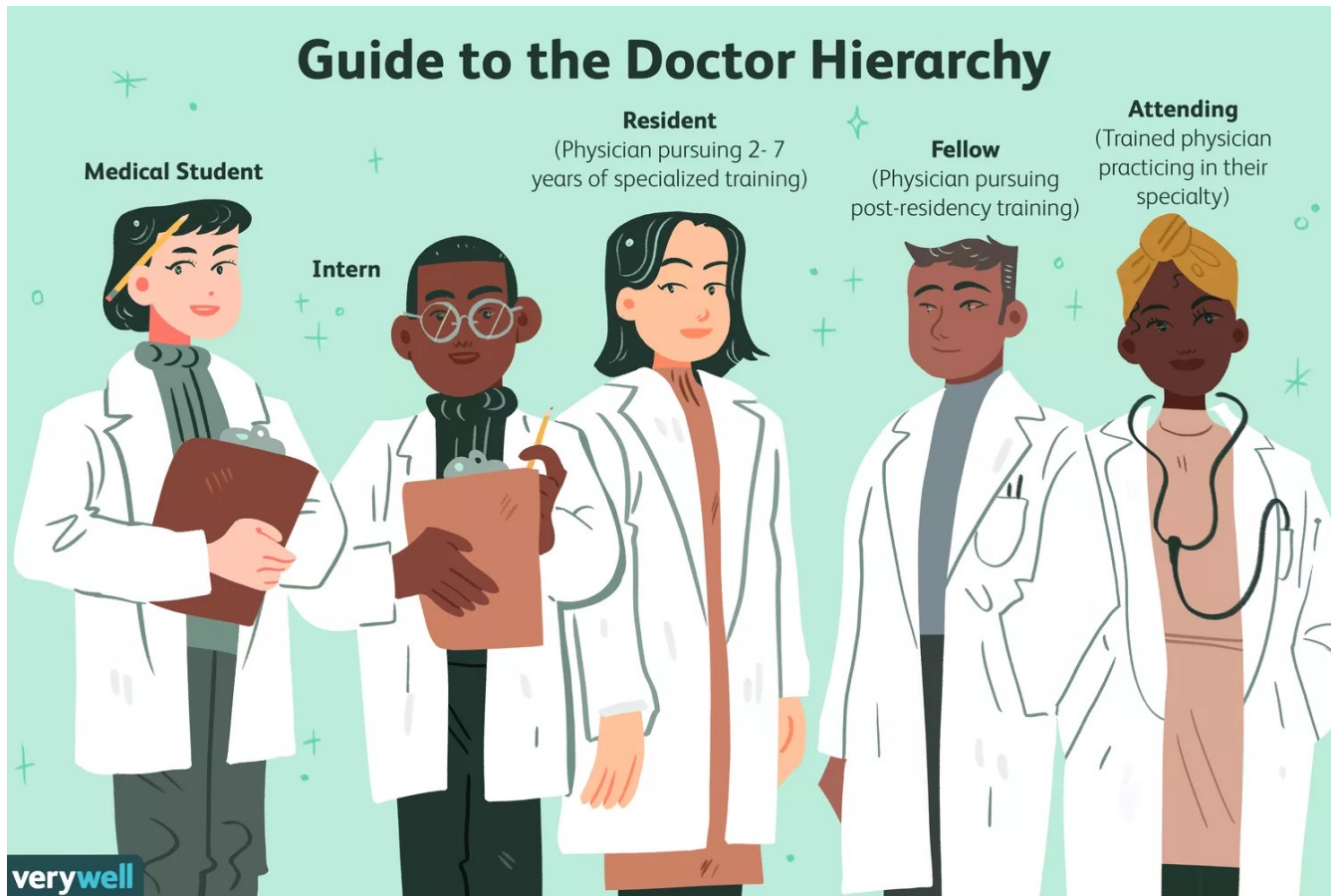
Outline

- Documentation in the clinical workflow
- Machine learning for clinical text
- Latent Dirichlet allocation for text data
- Neural language models for healthcare data
 - An evaluation of GPT-3 for information extraction from text data

Text data in clinical care

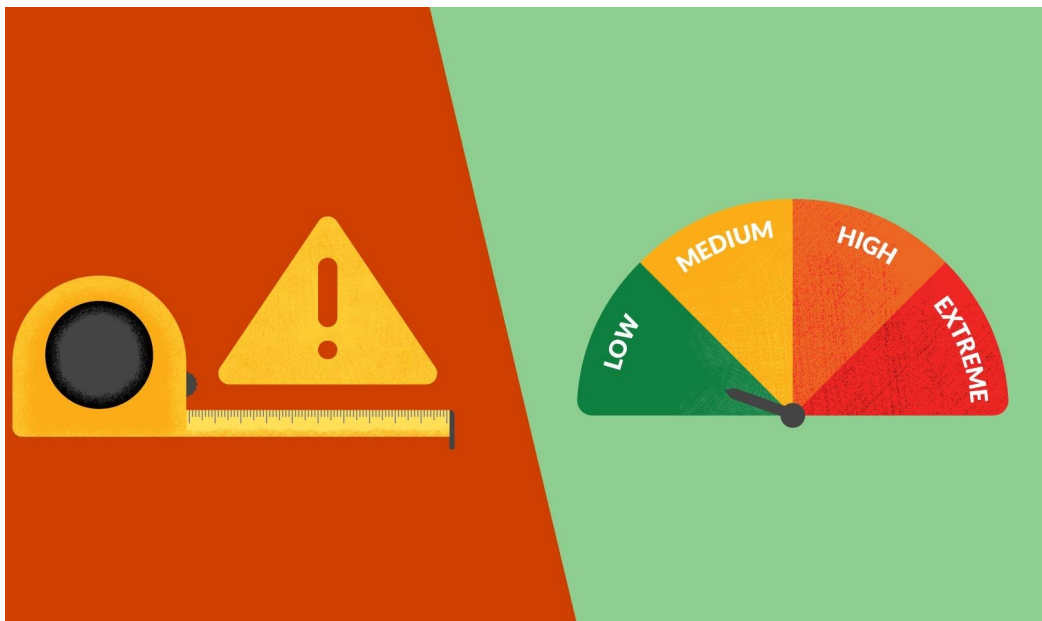
- Text data is ubiquitous in healthcare
- Most patient-clinician interactions often come with documentation
- **Why?**

Reason 1: One visit, many doctors



Vital to have clear documentation of healthcare needs for other personnel to take care of.

Reason 2: Risk management



- Errors in medicine can occur due to adverse events, unforeseen circumstances or negligence,
- Documentation serves to ensure that institutions have legal defensibility for actions taken

Reason 3 & 4: Tracking metrics & reimbursements

- Operational metrics:
 - How many lab-tests are we ordering per day?
 - What are the demographic details of our patient population, how are they changing?
 - How severe is the current case load in the ICU
- Reimbursement:
 - Records serve as back up for claims submitted for reimbursement

Clinical notes are a rich source of information about a patient

Mr. Blind is a 79-year-old white male with a history of diabetes mellitus, inferior myocardial infarction, who underwent open repair of his increased diverticulum November 13th at Sephsandpot Center.

The patient developed hematemesis November 15th and was intubated for respiratory distress. He was transferred to the Valtawnprinceel Community Memorial Hospital for endoscopy and esophagoscopy on the 16th of November which showed a 2 cm linear tear of the esophagus at 30 to 32 cm. The patient's hematocrit was stable and he was given no further intervention.

The patient attempted a gastrografin swallow on the 21st, but was unable to cooperate with probable aspiration. The patient also had been receiving generous intravenous hydration during the period for which he was NPO for his esophageal tear and intravenous Lasix for a question of pulmonary congestion.

On the morning of the 22nd the patient developed tachypnea with a chest X-ray showing a question of congestive heart failure. A medical consult was obtained at the Valtawnprinceel Community Memorial Hospital. The patient was given intravenous Lasix.

Slide credit: Peter Szlovits

Goals from text data

- There are a rich variety of tasks that we can leverage machine learning models for
 - Mapping text to medical ontologies
 - “Diabetes” -> E11
 - Identify if text contains protected information
 - Challenging edge cases: “Parkinson’s disease”
 - Identifying interventions
 - Which drug did the patient respond/not-respond to?
 - Summarization
 - What is a meaningful summary of patient state?

This lecture:

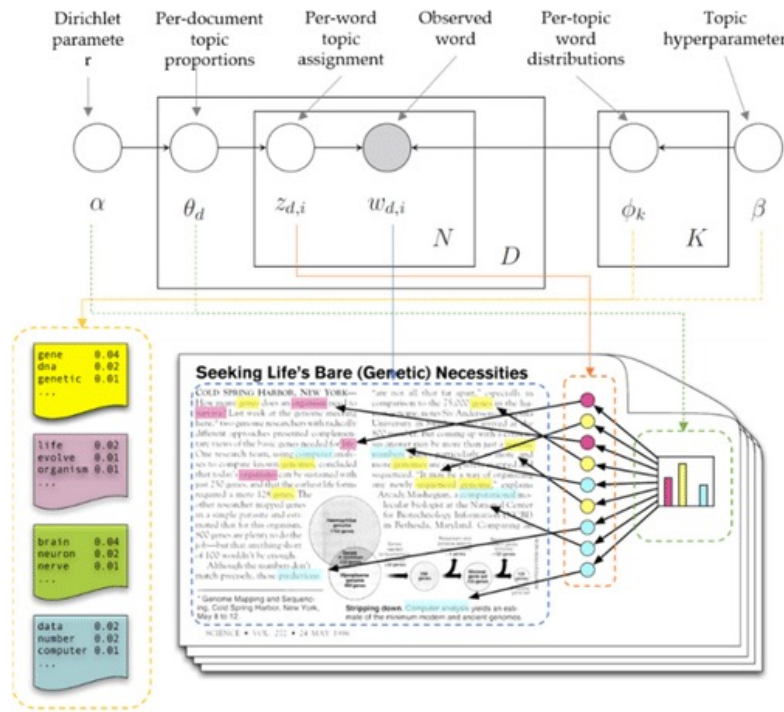
- Supervised learning:
 - Predicting outcomes from clinical notes
- Unsupervised learning:
 - Learning “representations” of patients

Supervised learning with text-data

- [Predicting early psychiatric readmission with natural language processing of narrative discharge summaries, Rumshisky et. al , Translational Psychiatry, 2016](#)
- Goal: 30-day re-admission
- Cohort: Patients in a psych inpatient ward between 1994-2012 with a diagnosis of major depression
- Baseline:
 - Support vector machine on clinical features + ~1000 most common words in notes
- ML model:
 - Latent Dirichlet Allocation
 - One of the most popular latent variable models of text data

Latent Dirichlet Allocation

(a) LDA document generation process



(b) An illustrative example of LDA document generation process

(c) Two outputs of LDA

(c-1) Per-document topic proportions (θ_d)

	Topic 1	Topic 2	Topic 3	...	Topic K
Doc 1	0.20	0.50	0.10	...	0.10
Doc 2	0.50	0.02	0.01	...	0.40
Doc 3	0.05	0.12	0.48	...	0.15
...
Doc N	0.14	0.25	0.33	...	0.14

(c-2) Per-topic word distributions (ϕ_k)

	Topic 1	Topic 2	Topic 3	...	Topic K
word 1	0.01	0.05	0.05	...	0.10
word 2	0.02	0.02	0.01	...	0.03
word 3	0.05	0.12	0.08	...	0.02
...
word N	0.04	0.01	0.03	...	0.07

Image credit: Insider Threat Detection Based on User Behavior Modeling and Anomaly Detection Algorithms, Kim et. al, Applied Sciences 2019

Table 2. Example topics for MDD patients readmitted with a psychiatric diagnosis within 30 days

<i>Terms</i>	<i>Topic annotation</i>
*patient alcohol withdrawal depression drinking end ativan etoh drinks medications clinic inpatient diagnosis days hospital <substance use treatment program name> use abuse problem number	Alcohol
*mg daily discharge anxiety klonopin seroquel clonazepam admission wellbutrin given md lexapro date b signed night low admitted sustained hospitalization	Anxiety
*ideation suicidal mood decreased hallucinations history depressed depression thought psychiatric energy denied sleep auditory appetite homicidal symptoms increased speech thoughts	Suicidality
*ect depression treatment treatments dr mg course <ECT physician name> symptoms received medications prior improved decreased medication md trials tsh continued qhs	ECT
*weight eating admission discharge hospital intake loss date hospitalization day dr week physical months prozac food increased md did anorexia	Anorexia
*seizure seizures intact eeg neurology normal temporal dilantin head bilaterally events activity weakness sensation disorder tongue neurologist brain loss tegretol	Seizure
*therapist mother program father disorder age school parents brother abuse treatment relationship outpatient college behavior partial plan currently group personality	Psychotherapy
*psychiatry suicide overdose attempt transferred depression transfer level tylenol hospital service unit normal floor screen tox room admission medical general	Overdose
*baby delivery bleeding vaginal breast feeding cesarean weight ibuprofen maternal newborn available p fever pregnancy sex estimated danger gp	Postpartum
*psychotic thought features paranoid psychosis paranoia symptoms psychiatric dose continued treatment mental cognitive memory risperidone people th somewhat interview affect	Psychosis

Abbreviation: MDD, major depressive disorder; ECT, electroconvulsive therapy.

Table 3. Comparison of models with and without inclusion of LDA topics

<i>Configuration</i>	<i>AUC</i>	<i>Sensitivity</i>	<i>Specificity</i>
Baseline = age/gender/insurance/ Charlson	0.618	0.979	0.104
Baseline+top-1 words	0.654	—	—
Baseline+top-10 words	0.676	—	—
Baseline+top-100 words	0.682	—	—
Baseline+top-1000 words	0.682	0.213	0.945
Baseline+75 topics (no words)	0.784	0.752	0.634

Abbreviation: AUC, area under the curve; LDA, Latent Dirichlet Allocation.

A (brief) history of unsupervised language models

- From bag-of-words to structural assumptions
- **What is it:** Language model is a statistical model of natural language text
- **How is it trained:** By maximizing the likelihood of a word/sentence

The dog jumped over the creek.

w_1 w_2 w_3 w_4 w_5 w_5 .

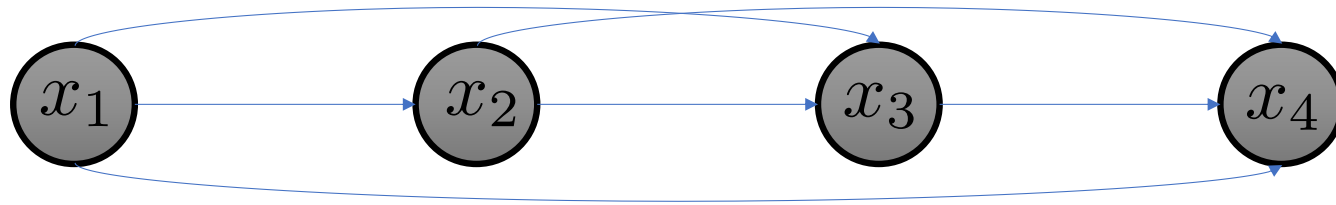
Each w_i is a word in a vocabulary set $[1.....V]$

Goal: Maximize $P(w_1....w_5)$

Language models over the years

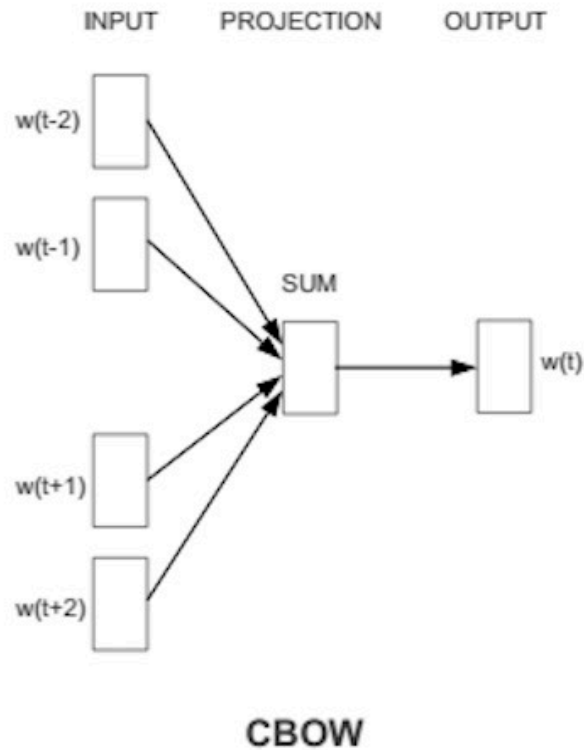
- The pre-deep learning era
 - **Ethos:** Need to have models that capture fine-grained structure in sentences
 - Parse trees
 - N-gram language models
 - Works well when model is not mis-specified but brittle when sentence syntax deviates from training data
- The post-deep learning era
 - **Ethos:** The context of a word is sufficient to predict the word
 - Word2Vec, Recurrent neural networks

K-gram models



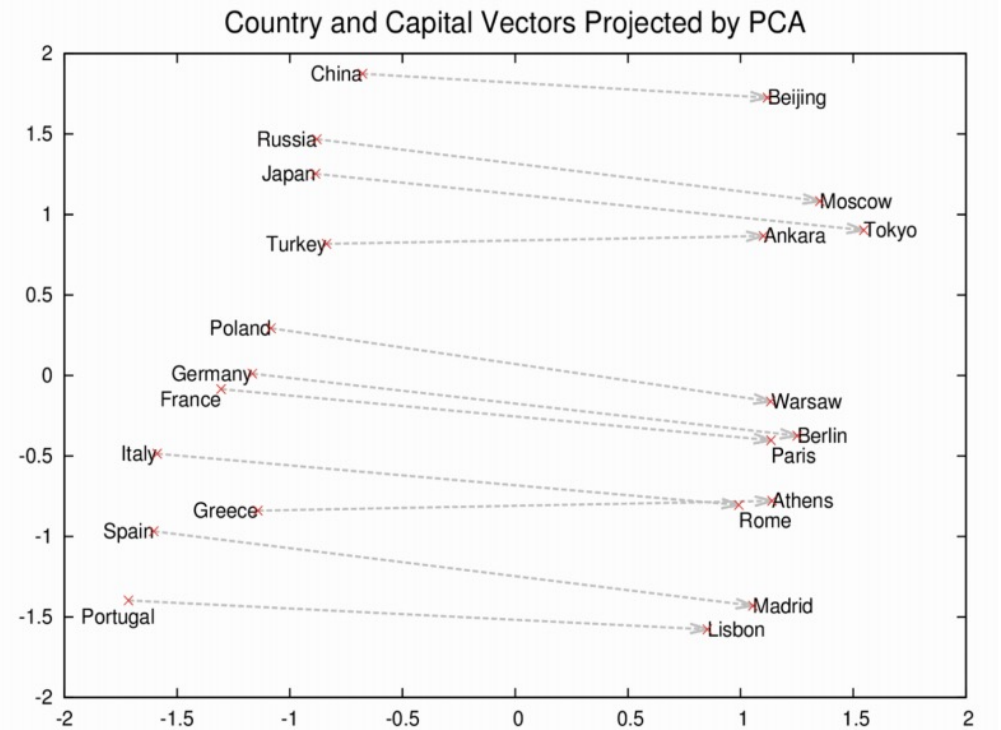
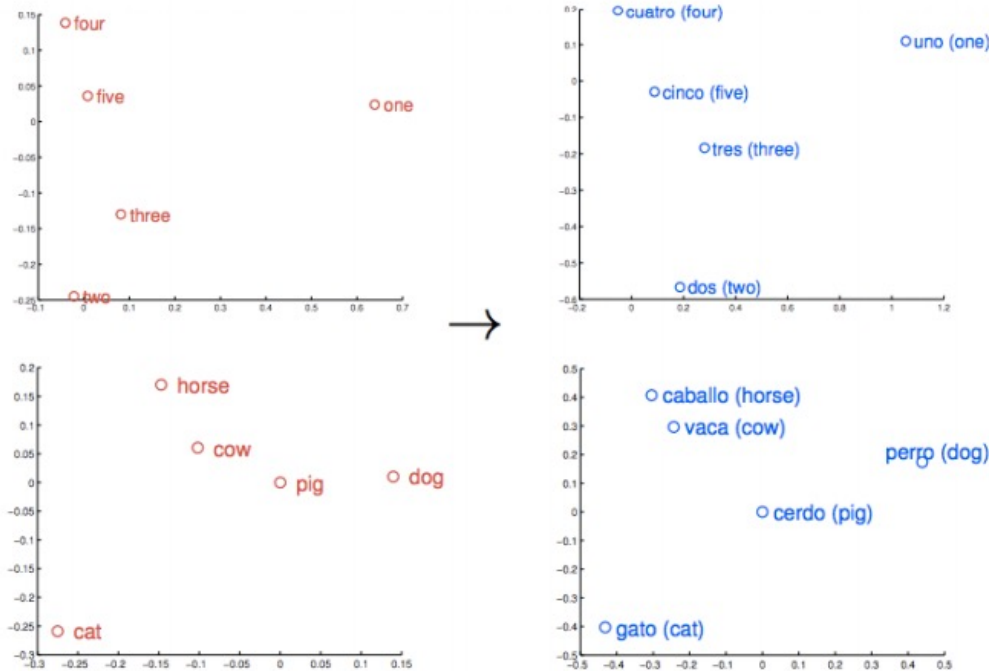
$$p(x_1, x_2, \dots, x_4) = p(x_1)p(x_2|x_1)p(x_3|x_{1..2})p(x_4|x_{1..,3})$$

Starting simple – word2vec



- Use a simple log-linear layer to predict the word given its context
- “You are who you spend your time with”

Word vectors capture interesting properties



Word2Vec in healthcare

- Clinical Concept Embeddings Learned from Massive Sources of Multimodal Medical Data, Beam et. al, PSB 2020
- “insurance claims database of 60 million members, a collection of 20 million clinical notes, and 1.7 million full text biomedical journal articles can be combined to embed concepts into a common space, resulting in the largest ever set of embeddings for 108,477 medical concepts”

Results – cui2vec

Table 1:

Comparison of *GloVe*, *PCA*, and *word2vec* for an embedding dimension of 500. Columns 1–4 report power to detect known relationships and column 5 reports the Spearman correlation between human assessments of concept similarity and cosine similarity from the embeddings. The best result for each each benchmark/dataset combination is shown in bold. The claims dataset contained only diagnosis codes and no drugs and so did not report results for the NDFRT benchmark.

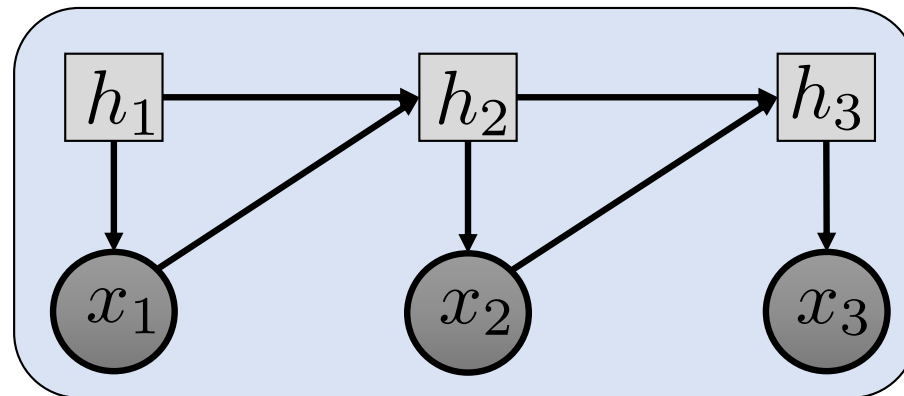
Data Source	Algorithm	Causative	Comorbidity	Semantic Type	NDFRT	Human Assessment
Claims	GloVe	0.56	0.73	0.29	-	0.45
	PCA	0.40	0.15	0.32	-	0.19
	word2vec (SVD)	0.54	0.50	0.40	-	0.45
PMC Articles	GloVe	0.59	0.57	0.28	0.54	0.60
	PCA	0.30	0.24	0.24	0.29	0.29
	word2vec (SVD)	0.83	0.59	0.49	0.84	0.67
	word2vec (original)	0.75	0.51	0.48	0.74	0.59
Clinical Notes	GloVe	0.39	0.73	0.51	0.11	0.34
	PCA	0.36	0.31	0.47	0.14	0.53
	word2vec (SVD)	0.75	0.52	0.74	0.49	0.59
Combined Data	GloVe	0.40	0.80	0.37	0.50	0.39
	PCA	0.24	0.23	0.30	0.37	0.47
	word2vec (SVD)	0.46	0.52	0.53	0.57	0.47

Limitation: only work based off the context around the word

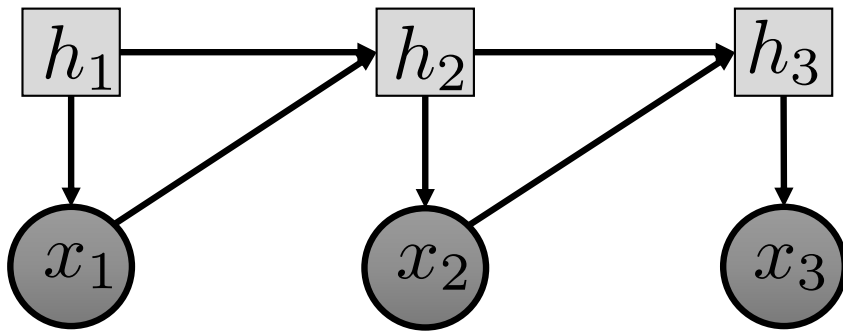
Recurrent Neural Networks

- Auto-regressive sequential models of data
- Forward recurrent neural network model
 - Each **hidden state** summarizes all the variables in the **past**

$$p(x_1, x_2, x_3) = p(x_1|h_1)\hat{p}(h_2|h_1)p(x_2|h_2)\hat{p}(h_3|h_2)p(x_3|h_3)$$



Recurrent neural networks in action

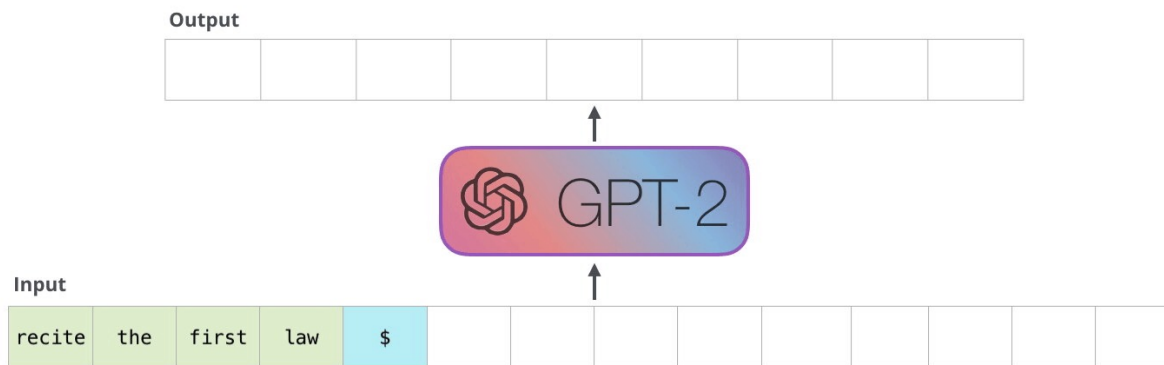


- Widely used for language modeling
- The parameterization of the functions that control how h behaves dictate the type of recurrent neural networks:
 - Long short-term memory (LSTM)
 - Gated recurrent units (GRU)

All of the expressive power in a RNN depends on the kinds of information that can be encoded in the hidden representation.

From RNNs to transformers with neural attention

- Recurrent neural networks are one parameterization of modeling auto-regressive distributions over long-sequences
- Transformers are a more powerful neural network to capture autoregressive distributions over the next word



Idea 1: Positional embeddings

- Intuition: The representation for a word should differ based on its location in a sentence

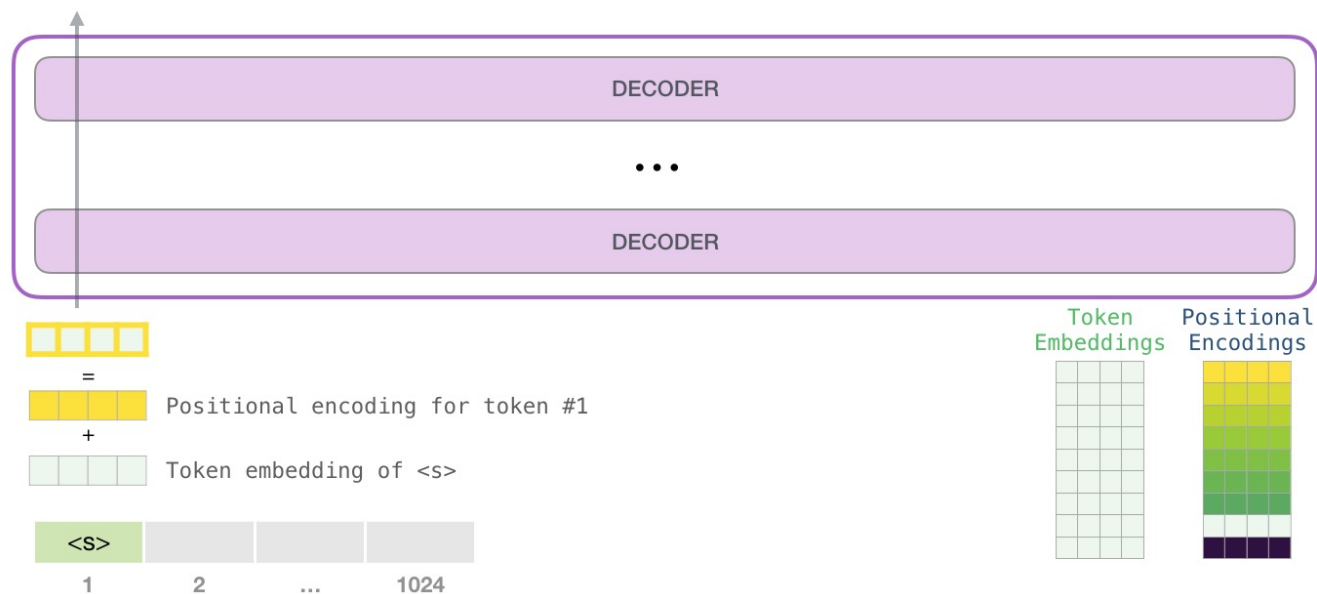





















Image credit: <https://jalammar.github.io/illustrated-gpt2/>

Idea 2: Self attention

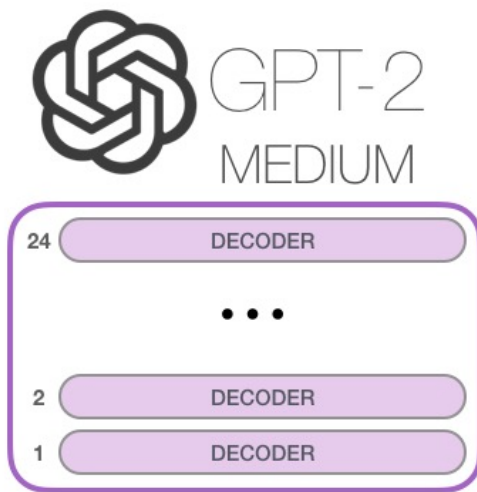
- The representation of a word should depend on the context that best describes it
- “it” refers to the the robot, we’d like the model to automatically learn that the words most related are “a, robot”

Word	Value vector	Score	Value X Score
<s>		0.001	
a		0.3	
robot		0.5	
must		0.002	
obey		0.001	
the		0.0003	
orders		0.005	
given		0.002	
it		0.19	
		Sum:	

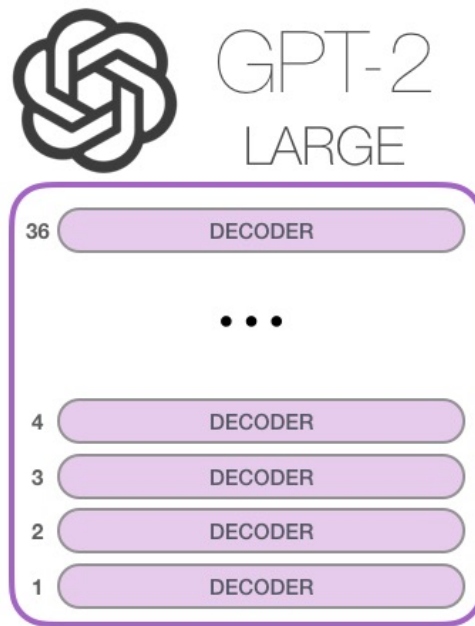
At the end of the month, we'll hear from your colleagues on the technical details of transformers and why they work



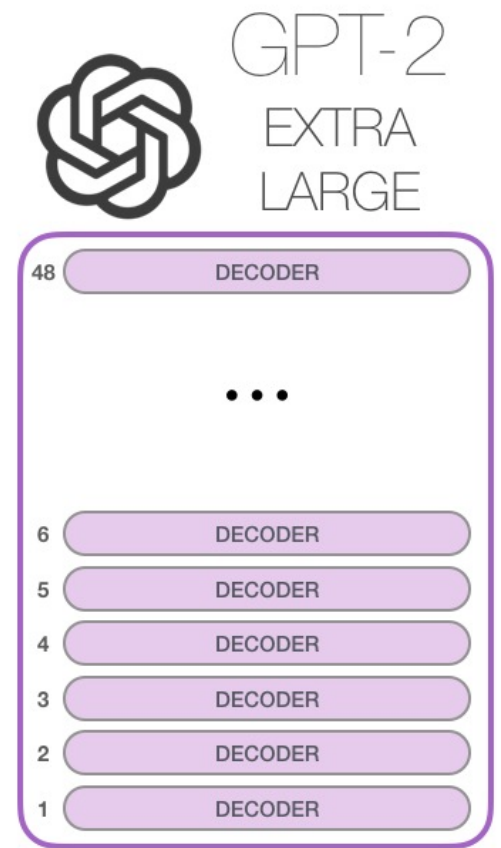
Model Dimensionality: 768



Model Dimensionality: 1024



Model Dimensionality: 1280



Model Dimensionality: 1600

GPT3

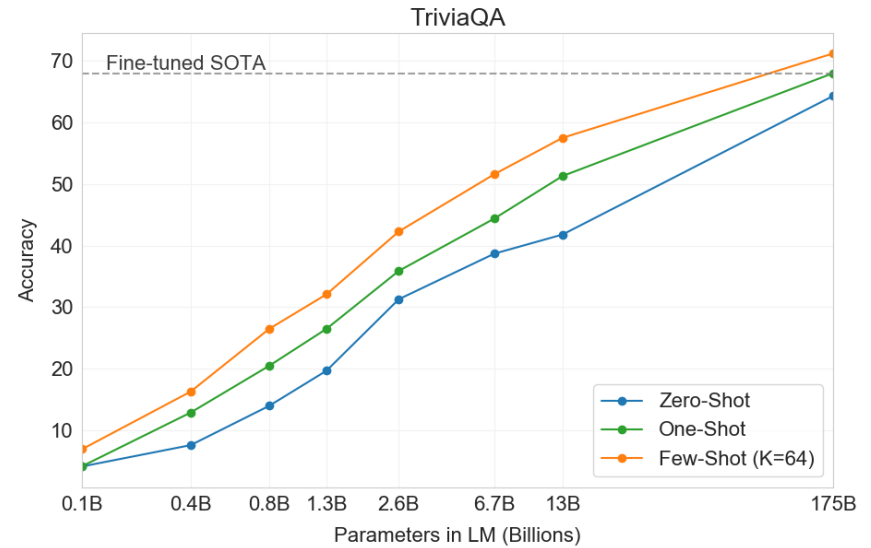
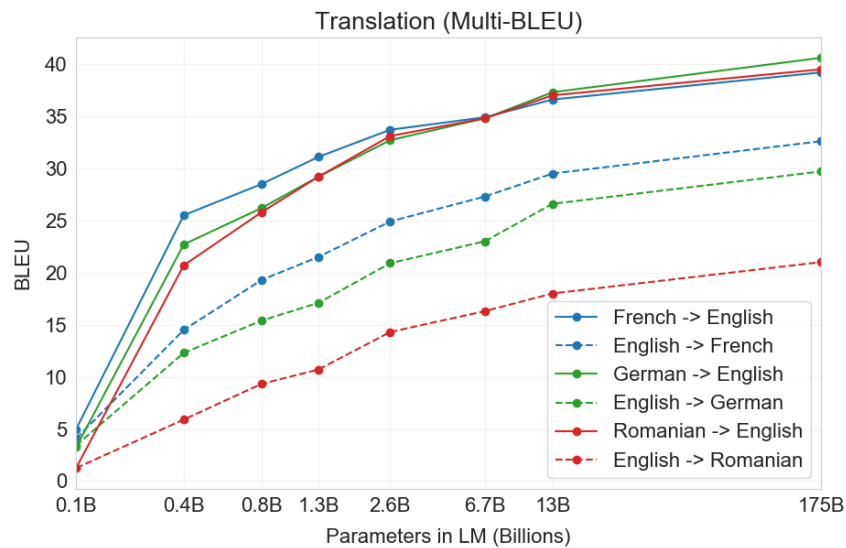
- Language Models are Few-Shot Learners, Brown et. al
 - 3 key ingredients
 - Leverages recent advances of deep learning (such as transformers)
 - Scales up the models to be [very] overparameterized
 - Trains on very very large datasets

Model Name	n_{params}	n_{layers}	d_{model}	n_{heads}	d_{head}	Batch Size	Learning Rate
GPT-3 Small	125M	12	768	12	64	0.5M	6.0×10^{-4}
GPT-3 Medium	350M	24	1024	16	64	0.5M	3.0×10^{-4}
GPT-3 Large	760M	24	1536	16	96	0.5M	2.5×10^{-4}
GPT-3 XL	1.3B	24	2048	24	128	1M	2.0×10^{-4}
GPT-3 2.7B	2.7B	32	2560	32	80	1M	1.6×10^{-4}
GPT-3 6.7B	6.7B	32	4096	32	128	2M	1.2×10^{-4}
GPT-3 13B	13.0B	40	5140	40	128	2M	1.0×10^{-4}
GPT-3 175B or "GPT-3"	175.0B	96	12288	96	128	3.2M	0.6×10^{-4}

Table 2.1: Sizes, architectures, and learning hyper-parameters (batch size in tokens and learning rate) of the models which we trained. All models were trained for a total of 300 billion tokens.

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

One model; many tasks; no supervision



A foray into GPT3 for clinical tasks

Warning

Upcoming results very are preliminary.
Not validated by a clinician.

Feedback, questions and concerns appreciated.

Language models for clinical question answering

Task 1

What does GPT 3 know about oncology?

- **Q: What is cancer immunotherapy?**
- **A:** Cancer immunotherapy is a treatment that uses the body's immune system to fight cancer. It is a type of biologic therapy.
- **Q: What is the life expectancy of patients suffering from small cell lung cancer?**
- **A:** The life expectancy of patients suffering from small cell lung cancer is about 12 months.
- **Q: What role does genetics play in the progression of cancer?**
- **A:** Genetics play a role in the progression of cancer.

Task 1

What does GPT 3 know about oncology?

- **Q: What is the BRCA1 gene?**
- **A: The BRCA1 gene is a gene that is associated with breast cancer.**
- **Q: How does the BRCA1 gene affect likelihood of breast cancer?**
- **A: The BRCA1 gene increases the likelihood of breast cancer.**
- **Q: How does SOD1 affect the likelihood of breast cancer?**
- **A: The SOD1 gene increases the likelihood of breast cancer.**
- **Q: What diseases are more likely due to mutations in the SOD1 gene?**
- **A: Diseases that are more likely due to mutations in the SOD1 gene include amyotrophic lateral sclerosis (ALS) and frontotemporal dementia.**

Task 1 [increase specificity wrt cancer]

What does GPT 3 know about multiple myeloma?

- **How to treat multiple myeloma?**

The treatment of multiple myeloma depends on the stage of the disease. The main goal of treatment is to reduce the number of abnormal plasma cells in the bone marrow.

Treatment may include:

Chemotherapy – chemotherapy is used to destroy cancer cells and slow down their growth. It may be used alone or in combination with other treatments.

– chemotherapy is used to destroy cancer cells and slow down their growth. It may be used alone or in combination with other treatments. Radiation therapy – radiation therapy uses high-energy rays to destroy cancer cells.

– radiation therapy uses high-energy rays to destroy cancer cells. Immunotherapy – this is a treatment that uses the patient's immune system to fight cancer cells.

– this is a treatment that uses the patient's immune system to fight cancer cells. Stem cell transplant – this treatment involves replacing the patient's bone marrow with healthy donor bone marrow.

– this treatment involves replacing the patient's bone marrow with healthy donor bone marrow. Targeted therapy – this treatment uses drugs that target specific molecules in cancer cells.

– this treatment uses drugs that target specific molecules in cancer cells. Surgery – surgery may be used to remove cancerous tissue.

– surgery may be used to remove cancerous tissue. Supportive care – supportive care helps patients feel better and live longer. It includes: Pain relief Control of nausea and vomiting Help with breathing Control of anemia Control of bleeding

Task 1 [increase specificity wrt cancer]

What does GPT 3 know about multiple myeloma?

- **What are the possible complications of multiple myeloma?**

Complications of multiple myeloma include:

Bone fractures – the weakened bones may break easily.

– the weakened bones may break easily. Infection – patients with multiple myeloma are at risk of developing infections.

– patients with multiple myeloma are at risk of developing infections. Anemia – the bone marrow produces fewer red blood cells, which carry oxygen to the body's tissues.

– the bone marrow produces fewer red blood cells, which carry oxygen to the body's tissues. Thrombosis – blood clots may form in the veins of patients with multiple myeloma.

– blood clots may form in the veins of patients with multiple myeloma. Stroke – patients with multiple myeloma are at risk of having a stroke.

– patients with multiple myeloma are at risk of having a stroke. Heart disease – patients with multiple myeloma may develop heart disease.

– patients with multiple myeloma may develop heart disease. Infection – patients with multiple myeloma are at risk of developing infections.

– patients with multiple myeloma are at risk of developing infections. Anemia – the bone marrow produces fewer red blood cells, which carry oxygen to the body's tissues.

– the bone marrow produces fewer red blood cells, which carry oxygen to the body's tissues. Thrombosis – blood clots may form in the veins of patients with multiple myeloma.

Task 1 [ask about treatments]

What does GPT 3 know about multiple myeloma?

- **How does bortezomib treat multiple myeloma?**

Bortezomib is a proteasome inhibitor. The proteasome is an enzyme that breaks down proteins in the body. Bortezomib stops the proteasome from breaking down proteins, which allows the immune system to attack cancer cells.

Bortezomib is used to treat multiple myeloma that has returned (recurred) or does not respond to other treatments.

Bortezomib is usually given after other treatments have been tried without success. It may be used before a stem cell transplant.

Task 1 [ask about clinical protocol]

What does GPT 3 know about multiple myeloma?

- **How to doctors track multiple myeloma?**

The most common test used to monitor multiple myeloma is the serum free light chain assay. This test measures the amount of free light chains in the blood. The higher the level of free light chains, the more active the disease.

Other tests that may be done include:

Complete blood count (CBC)

Blood chemistry tests

Bone marrow aspiration and biopsy

X-rays, CT scans, and MRI scans of the bones and soft tissues (to look for bone lesions)

Task 1 [**Failure case**]

What does GPT 3 know about multiple myeloma?

- **What are genetic markers of individuals who respond to treatment for multiple myeloma?**

What are the genetic markers of individuals who respond to treatment for multiple myeloma?

What are the genetic markers of individuals who respond to treatment for multiple myeloma?

Language models for information extraction

Task 2 – Extract structured information from unstructured data

- Neuro: Patient intermittently awakens and moves head from side to side. PERRL. Grimaces with painful stimuli. remains on fentanyl at 250mcg/hr and was started on ativan gtt at 2mg/hr. Has bolused patient with extra ativan X2 for agitated episodes. Team would like fentanyl slowly weaned and today's goal was 250mcg/hr which he has been tolerating well. It is notable that his agitation as demonstrated by increased resp effort, increased HR, and hypertension have worsened when stimulated by family. Family have been encouraged to decrease amount of 1:1 time with patient and to decrease stimulation of patient. Lung sounds are diminished at bases and coarse in upper fields. Have suctioned 3-4X this shift with scant yellow output. Was changed from AC to PC yesterday and has been tolerating this change well throughout today. Current settings are FiO2 40%/rate28/Pip34/Peep8/Drive pressure 26. ET tube is 8.5 and is 25 at the lips. Centrally located. SpO2 >94% throughout day. This am ABG with pH 7.39
GU/GI: Foley patent with dark amber urine. Urine output about 125cc/hr. This am BUN 61/ Cr 1.7 which is relatively unchanged from yesterday. As noted above, dulcolax supp. administered with good results. OG tube continues to drain brown drainage. Nasal/post-pyloric tube in place with peptamen at 10cc/hr. Abdomen remains very distended. Bowel sounds are hypoactive in all quadrants. Continues to receive TPN at 79cc/hr. On insulin gtt at 15 units/hr. Required regular insulin bolus and increased insulin gtt rate throughout the day, current fingerstick is 112. Monitor new bag of TPN for changes in dextrose or insulin.
- | fentanyl | 250mcg/hr |
- | ativan | 2mg/hr |
- | Urine output | 125cc/hr |
- | BUN | 61 |
- | Cr | 1.7 |
- | TPN | 79cc/hr |
- | Insulin | 15 units/hr |
- | TPN dextrose | 5.5% |
- | TPN insulin | Regular |
- | TPN dextrose | 5.5% |
- | TPN insulin | Regular |

Task 2 – Extract structured information from unstructured data

- Alert and oriented X3 throughout night. CT and MRI of head done yesterday following Cardiac Cath related to facial droop, slurred speech and right sided weakness. Radiological testing showed no acute bleed- ? embolic process. Speech continues to be slurred (though clearer as the night progresses) and there is a right sided facial droop noted. Right hand grasp weak- patient is able to raise and hold right arm up independently. No deficit noted to right arm. Able to raise and hold bilateral lower extremities and strength to lower extremities are equal with flexion and extension. PERRL. Cardiovascular: S/P cardiac cath with stent placement to RCA. Heart rate largely regular though occasionally irregular with cardiac monitor showing sinus brady to sinus rhythm with occasional APC's and brief runs of ventricular bigeminy. HR 56-74 throughout night. BP 141/39-174/42 with goal to maintain SBP >140 for cerebral perfusion. Have not started nitro gtt which team is aware of and would like to be called in the event it became necessary to start. Peripheral pulses are weakly palpable to bilateral lower extremities. Integrelin infusing at 1mcg/kg/min via IV. Respiratory: Initially on O2 at 3lpm. While sleeping patient's saturation level dropped to 89% so O2 increased to 6lpm via N/C and Cool mist mask added at 50% FiO2 with resulting SpO2 92-98%. Lung Sounds are diminished in bases and clear in upper fields. GU/GI: Foley in place with excellent output. Labs done at 4pm on indicate that Cr 1.5, BUN 27. Abdomen is soft/non-tender with normal active bowel sounds in all quadrants. Fingertick blood glucose 111 at 2200. HS NPH insulin held related to glucose level and NPO status. Skin Integrity: There is a small bruise evident to right groin and a dressing in place to left groin where sheath was removed last pm. There is a duoderm in place to coccyx. S/P amputation of 2nd toe of left foot See carevue for specifics related to bilateral lower extremity wound care. ID: in place to right anterior forearm and IVF infusing to left forearm. Afebrile throughout night. Continues on Levofloxacin and Metronidazol po. Plan: Patient remains a Full Code. AM labs pending. Patient to receive PT consult for question of extended care placement for rehabilitation.
- | HR | 56-74 |
 - | BP | 41/39-174/42 |
 - | O2 initial | 3lpm |
 - | O2 final | 6lpm |
 - | WBC | 9200 |
 - | Platelets | 100 |
 - | INR | 1.5 |
 - | BUN | 27 |
 - | Cr | 1.5 |
 - | Glucose | 111 |
 - | Lactate | 13.6 |

Language models for summarization

Summarizing patient note

- **General:** Patient appeared more comfortable today than in the past two. His fentanyl drip has been gradually decreased and an ativan drip was started today in the hopes to slowly wean off fentanyl. Three loose stools today after dulcolax (one spec sent to lab for C.Dif). Multipodis boot added to care. **Neuro:** Patient intermittently awakens and moves head from side to side. **PERRL.** Grimaces with painful stimuli. remains on fentanyl at 250mcg/hr and was started on ativan gtt at 2mg/hr. Has bolused patient with extra ativan X2 for agitated episodes. Team would like fentanyl slowly weaned and today's goal was 250mcg/hr which he has been tolerating well. It is notable that his agitation as demonstrated by increased resp effort, increased HR, and hypertension have worsened when stimulated by family. Family have been encouraged to decrease amount of 1:1 time with patient and to decrease stimulation of patient. **Cardiovascular:** Heart rate regular and Tachy with range 110's-130's. BP with better control today though episodically SBP in 180's-190's and DBP in 100's. Echo done yesterday shows small pericardial effusion which team reports will likely not need any intervention. Peripheral pulses by doppler only. CVP 9-12 today.

Patient is doing better today and is tolerating his meds well. He has been weaned off fentanyl and is now on ativan. **He is still agitated and has been given extra ativan to control this.**

Future work

- **Research questions:** Understanding how GPT-3 processes medical queries and stores information among the parameters
 - Is clinical language natural?
 - What has GPT-3 learned about human anatomy?
 - How can we improve the knowledge contained within the model?
- **Limitations:** GPT-3 not trained on a medical corpus
 - **Warning: Biases in the model are not well understood!**
 - **They do not process information in the same way that humans do**
 - **Expensive to train and sample from these models**
 - **Need more research to understand what they are and how they manifest**