COLLABORATIVE SEMI-SUPERVISED LEARNING FRAMEWORK FOR CLINICAL NLP

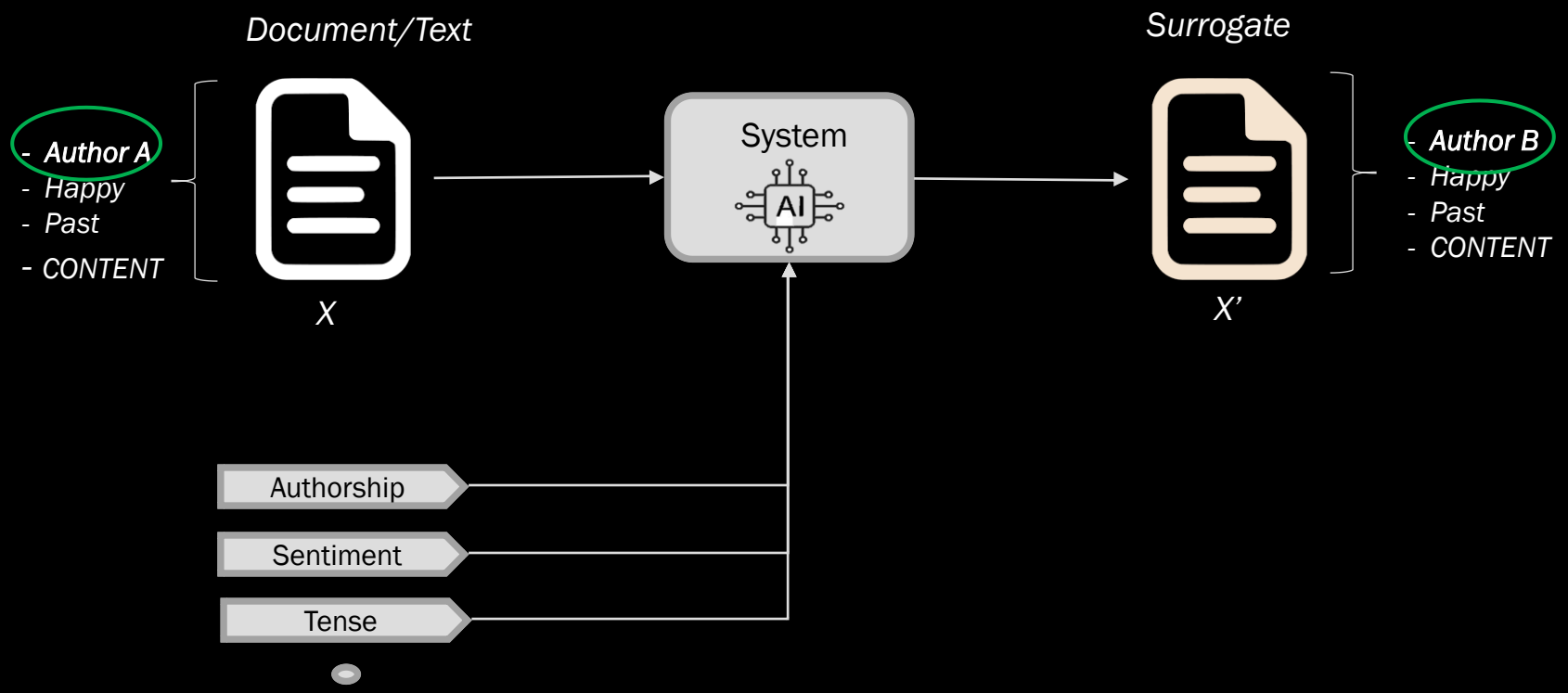
Master Thesis

Computer Science for Digital Media

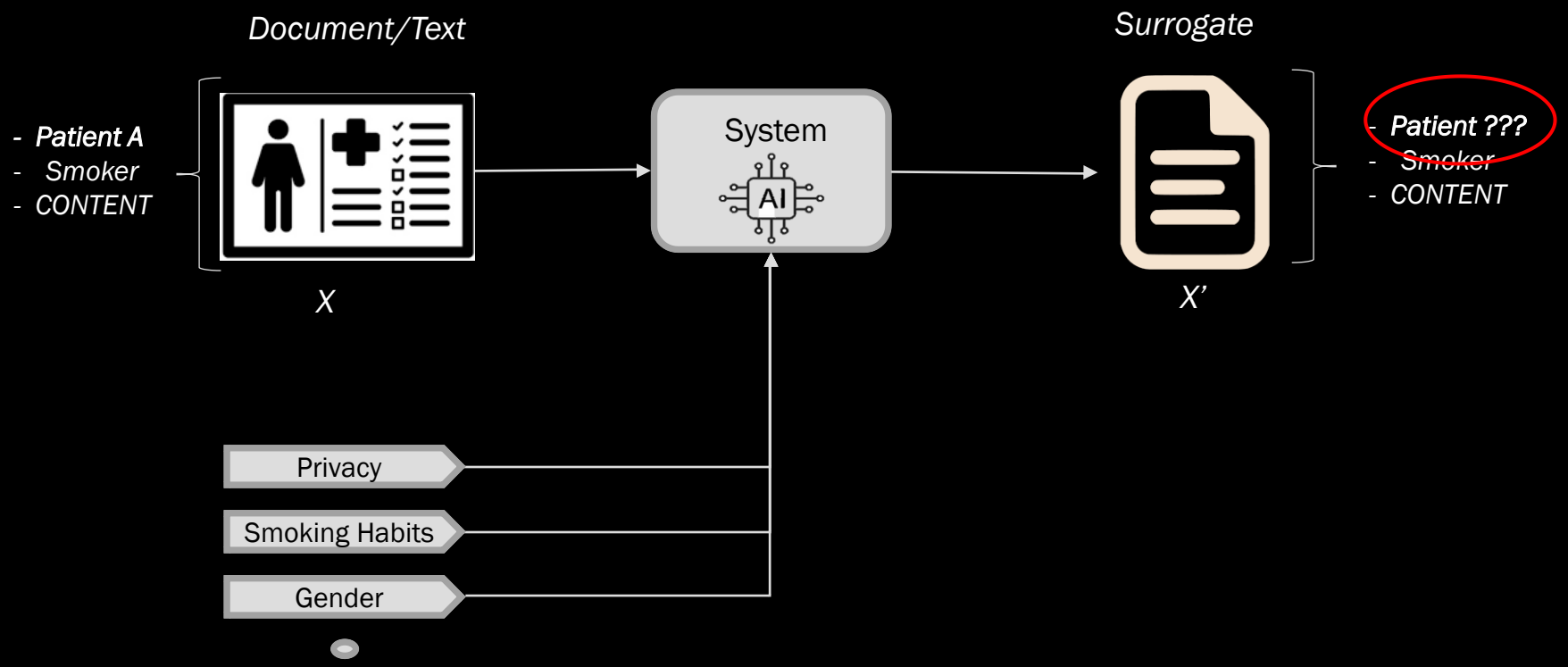
Bauhaus University Weimar 2022

SEBASTIAN LAVERDE ALFONSO

PROBLEM STATEMENT



PROBLEM STATEMENT



SAMPLE MEDICAL RECORDS

```
<RECORD ID="635">
<TEXT>
<PHI TYPE="ID">779810048</PHI>
<PHI TYPE="HOSPITAL">FIH</PHI>
<PHI TYPE="ID">8956861</PHI>
<PHI TYPE="ID">641681</PHI>
<PHI TYPE="ID">027815</PHI>
<PHI TYPE="DATE">12/02</PHI>/1998 12:00:00 AM
( 1 ) STATUS POST MOTOR VEHICLE COLLISION .
( 2 ) GRADE
Unsigned
DIS
Report Status :
Unsigned
DISCHARGE SUMMARY NAME :
<PHI TYPE="PATIENT">CHIRDVAIA , RITOC M</PHI>
UNIT NUMBER :
<PHI TYPE="ID">427-83-75</PHI>
ADMISSION DATE :
<PHI TYPE="DATE">12/02</PHI>/1998
DISCHARGE DATE :
<PHI TYPE="DATE">12/02</PHI>/1998
PRINCIPAL DIAGNOSIS :
( 1 ) Status post motor vehicle collision .
( 2 ) Grade 4 liver laceration through the left lobe of the liver and into the hilum and retrohepatic cava
( 3 ) Splenic laceration .
( 4 ) Severe left pulmonary contusion .
( 5 ) Cardiac contusion .
HISTORY OF PRESENT ILLNESS :
Mr. <PHI TYPE="PATIENT">Cuchsli</PHI> is a 17-year-old male who was brought via Medflight to the <PHI TYPE=
motor vehicle collision .
Per report , he hit a highway railing and subsequently a bridge embankment sustaining multiple injuries .
Per report , there was extensive damage to the vehicle and he required a very prolonged extrication time .
His accident occurred , per report , at 6:45 in the morning , and he was brought to the <PHI TYPE="HOSPITA
```

Sample from the *Automatic De-Identification Challenge*

→ PHI categories (3 out of 18): names of Hospitals, Doctors and Patients.

REAL RECORD CONTAINING SMOKING STATUS INFORMATION

,812367409SH21952193

06/19/1991 12:00:00 AM

Discharge SummarySignedDIS

Admission Date :06/19/1991

Report Status :Signed

Discharge Date :

Independence Day

HISTORY OF PRESENT ILLNESS :

The patient is an 84-year-old woman with a history of rheumatoid arthritis .She is now status post three myocardial infarctions .She has had progressive deformity and rheumatoid arthritis of her right knee .She presented at this time for a right total knee replacement .

PAST MEDICAL HISTORY :

As above .Appendectomy .Cholecystectomy .Left total knee replacement in 1977 .Pepticulcer disease .MEDICATIONS :On admission included Inderal , 40 mgpo q.i.d. ; Aldomet , 250 mg po t.i.d. ; apresoline , 250 mg po t.i.d. ; Nitropaste , one - half inch q p.m. ; Zantac , 150 mg po q p.m. ; Lasix , 20 mg po q day ; allopurinol , 300 mg po b.i.d. and Clinoril , 25 mgpo b.i.d.HABITS She does not smoke or drink

ALLERGIES :

ECOTRIN .MINIPRES .TAGAMET .HALDOL .

PHYSICAL EXAMINATION :

On admission revealed an elderly woman in no acute distress .Temperature 97.6 .Pulse 80 .Respiratory rate 18 .Blood pressure 200/86 .Skin was without rashes or breakdown .Lungs were clear to auscultation and percussion .Heart revealed a Grade IV / VI systolic ejection murmur .Abdomen was soft , nontender , no masses .Extremities revealed skin was intact to both lower extremities .Right leg demonstrated flexion from 5 degrees to 135 degrees .Left leg demonstrated flexion from 0 to 130 degrees .She had tenderness in her right knee .She had crepitus in her right knee .Sensory and motor function was intact .

LABORATORY DATA :

On admission included x-rays which demonstrated severe degenerative disease of her right knee .She had a creatinine of 2.0 .BUN 59 .Hematocrit of 31.4 .EKG revealed left bundle branch block with no acute ischemic changes .Urinalysis demonstrated a small amount of blood but no evidence of infection .

HOSPITAL COURSE :

Rheumatoid arthritis :The patient underwent a right total knee replacement after Cardiology clearance .She tolerated the procedure well , however , in the immediate postoperative period she developed confusion and adelirium status state .She was evaluated by Neurology and followed carefully .All of her pain medications were discontinued and she was maintained with a sitter .Psychiatry Service evaluated the patient and she was scheduled for a head CT .Head CT demonstrated no evidence of any stroke or acute compromise but there appeared to be chronic atrophy .Further evaluation consisted of a Urology consult which followed for her mildly elevated creatinine .The patient was followed by the Cardiology Service postoperatively and demonstrated no evidence of myocardial infarction in the immediate postoperative period .Note that she cleared mentally spontaneously over approximately seven days postoperatively .She subsequently did well with physical therapy .She was cleared for discharge after achieving flexion beyond 100 degrees .She was able to ambulate up and down stairs with crutches .She had x-rays taken confirming good alignment of her prosthesis .Ultrasound ruled out evidence of a deep venous thrombosis .She was therefore discontinued on Coumadin .

DISCHARGE DIAGNOSES :

STATUS POST RIGHT TOTAL KNEE REPLACEMENT .CONFUSION .CHRONIC RENAL FAILURE .

DISPOSITION :

The patient was discharged to home in satisfactory condition .

MEDICATIONS :

On discharge included Zantac , 150 mg po q p.m. ; Lasix , 20 mg po q day ; allopurinol , 300mg po b.i.d. ; Clinoril , 325 mg po b.i.d. ; Nitropatch , one - half inch q.p.m. ; apresoline , 250 mg po t.i.d. ; Aldomet , 250 mg po t.i.d. ; Inderal , 40 mg po q.i.d.IL799/1282RAMAG L. TROISQUARCKAYS , M.D. KK5D :07/11/91Batch :8122Report :H7634J2T :07/15/91Dictated By: HEAGLE , M.D. report_end] ed and ct scan on ct scan at consult but showed no active"

Sample from the *Automatic Classification Challenge*
Patient's label: Non-smoker

DE-IDENTIFICATION

Deleting Identifiers from the data that directly or indirectly point to a person (or entity).
Disentangling private aspects from medical records.

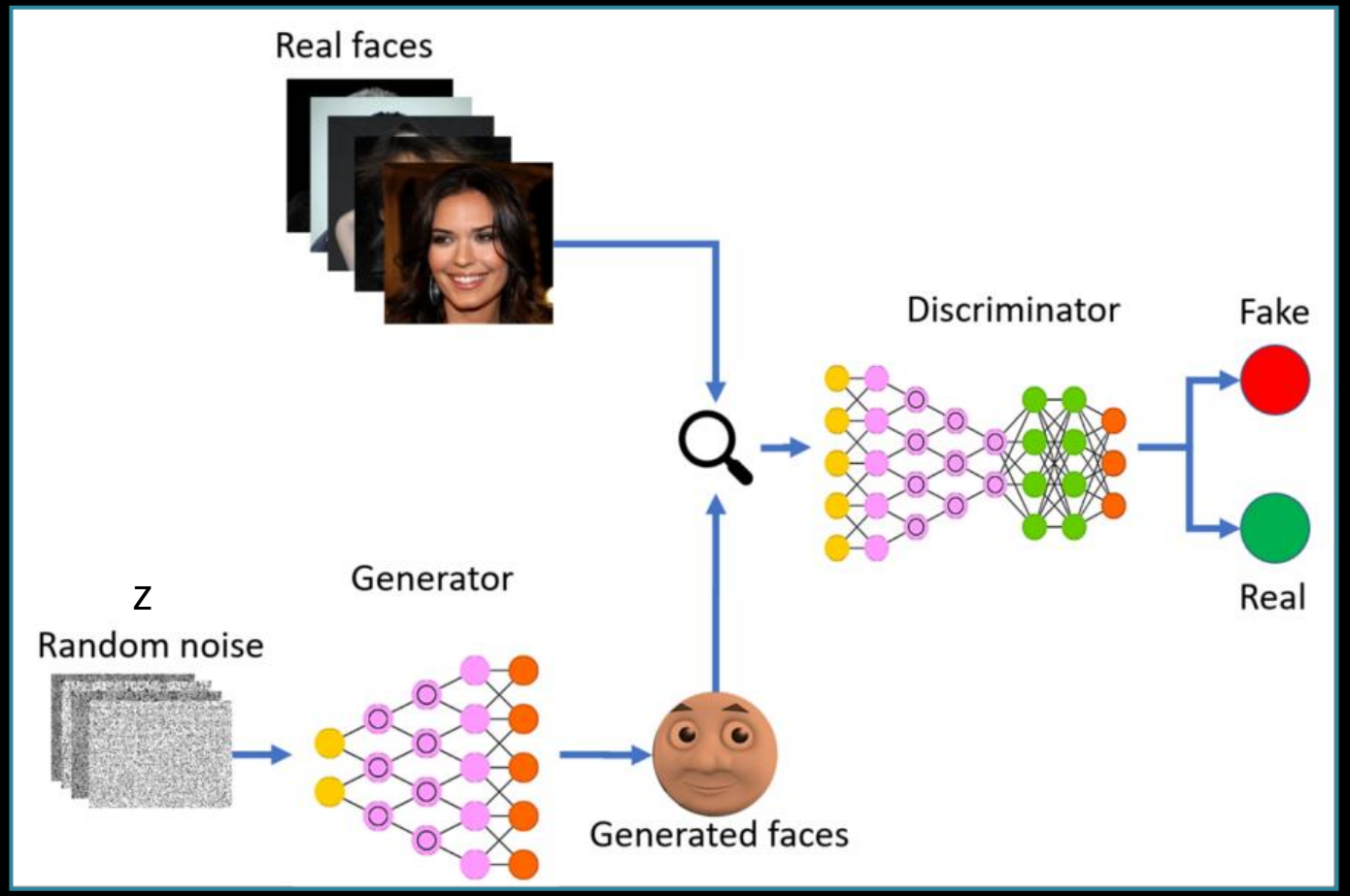
CLASSIFICATION

Automatically inferring a label about an aspect of the medical records

Privacy
Safe-to-Share Data

Utility
Useful Data

GENERATIVE MODELS
GENERATIVE ADVERSARIAL NETWORK (GAN)

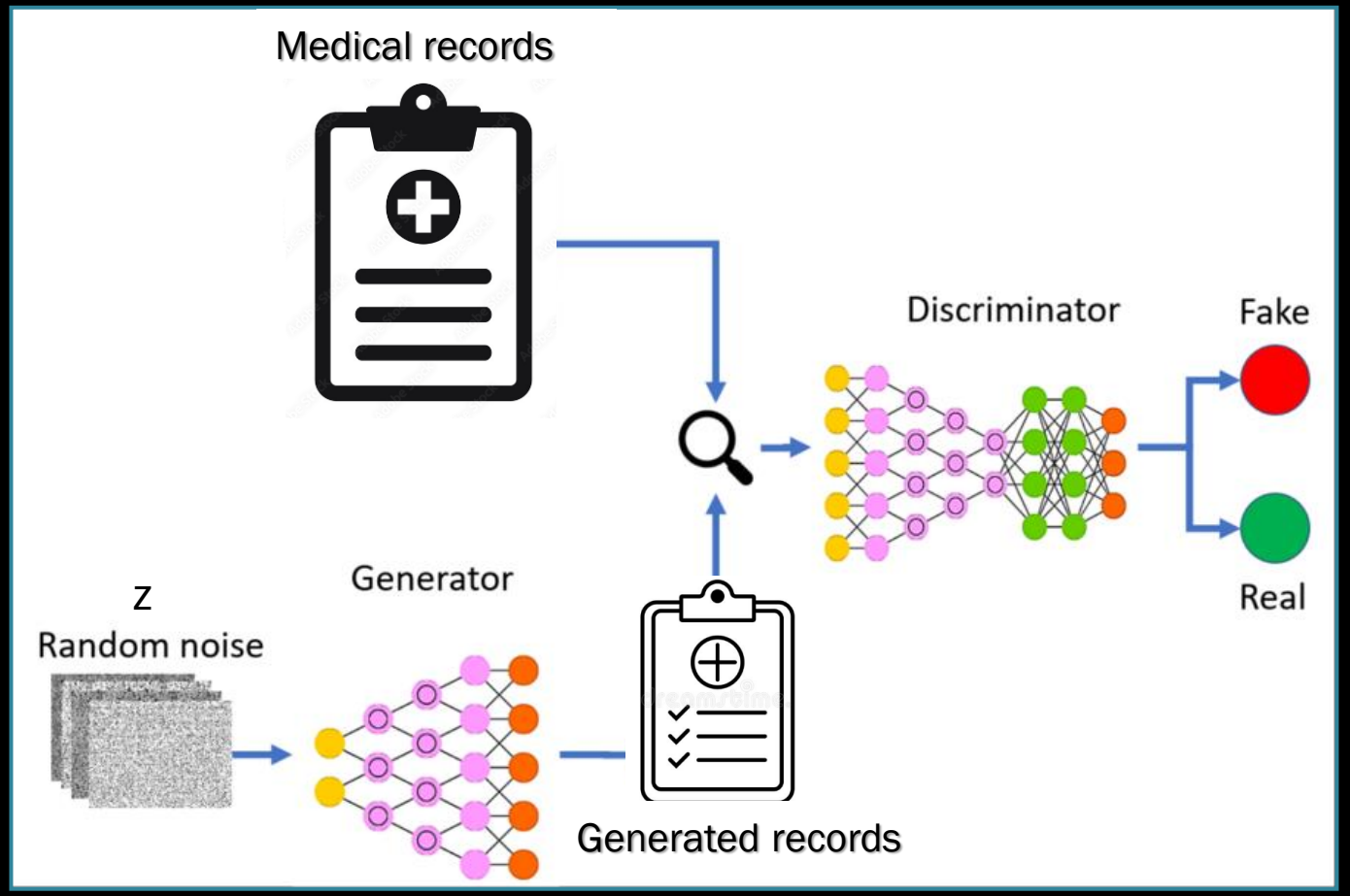


GAN for Face Generation

Taken from: https://www.kdnuggets.com/2020/03/generate-realistic-human-face-using-gan.html

Exploring the latent after training

GENERATIVE MODELS
GENERATIVE ADVERSARIAL NETWORK (GAN)

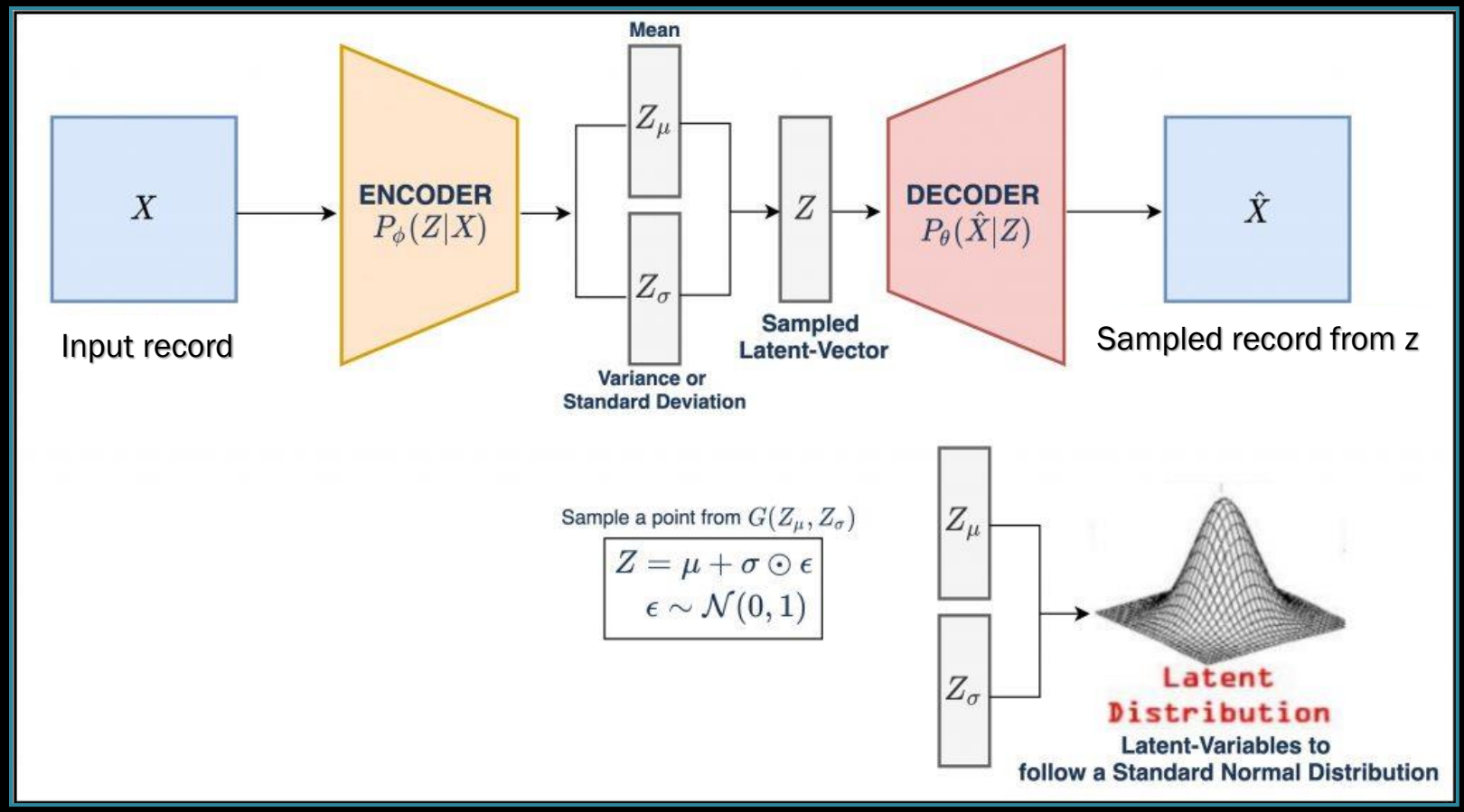


GAN for Text Generation

Modified from: https://www.kdnuggets.com/2020/03/generate-realistic-human-face-using-gan.html

Exploring the latent after training

GENERATIVE MODELS
VARIATIONAL AUTOENCODER (VAE)



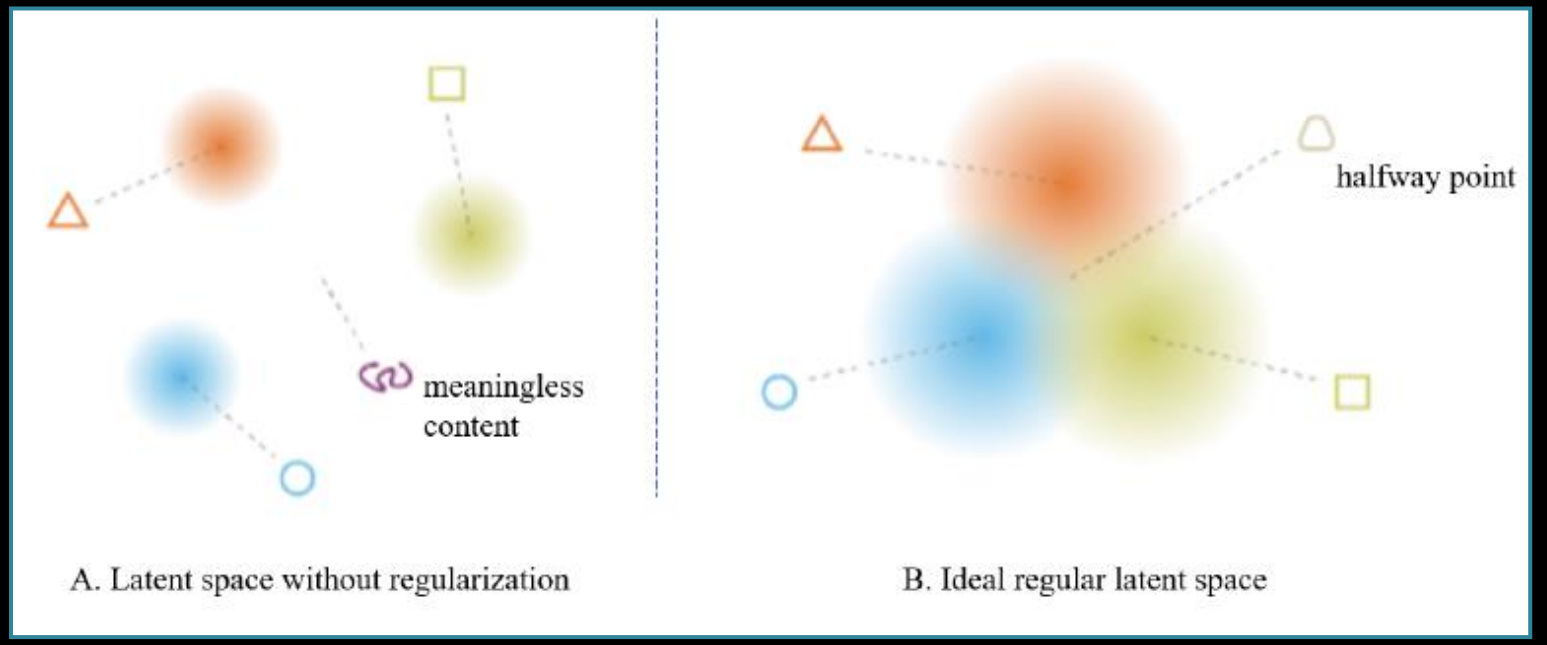
VAE architecture for generation; Latent space z

Taken from: https://learnopencv.com/variational-autoencoder-in-tensorflow/

LOSS
VARIATIONAL AUTOENCODER (VAE)

$$\text{loss} = \underbrace{\| x - \hat{x} \|^2}_{\text{Reconstruction Loss}} + \overbrace{KL[\, N(\mu_x, \sigma_x), N(0, I)\,]}^{\text{KL-Divergence Regularization}}$$

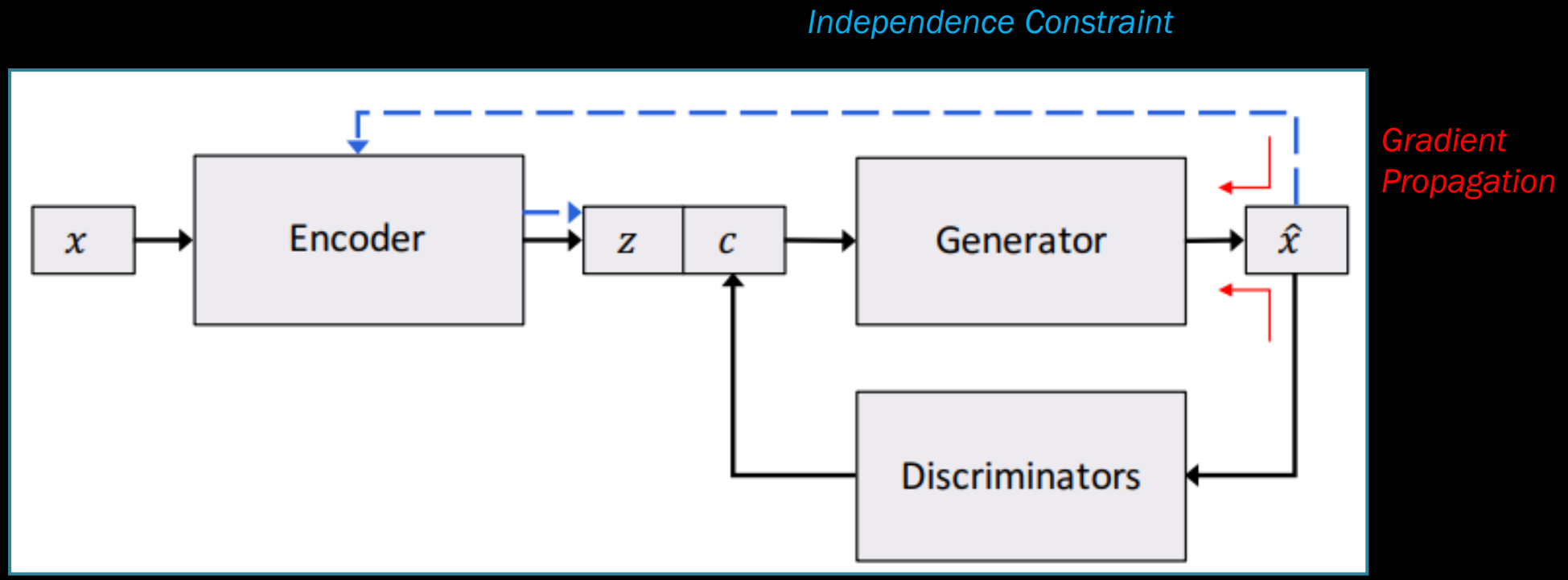
$\| x - d(z) \|^2$



Effect of regularization

Taken from: https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

CONTROLLED TEXT GENERATION
CTG-VAE



Original pipeline from base paper

Taken from: https://learnopencv.com/variational-autoencoder-in-tensorflow/

CONTROLLED TEXT GENERATION CTG-VAE

Generator Objective

$$\min_{\theta_G} \mathcal{L}_G = \mathcal{L}_{\text{VAE}} + \lambda_c \mathcal{L}_{\text{Attr},c} + \lambda_z \mathcal{L}_{\text{Attr},z},$$

Discriminator Objective

$$\min_{\theta_D} \mathcal{L}_D = \mathcal{L}_s + \lambda_u \mathcal{L}_u,$$

Algorithm: Text Generation Controlled by D_i

Inputs:

- Unlabeled text corpus $X = \{x\}$
- Labeled corpus $X_L = \{(x_L, c_L)\}$
- Value for balaning parameters λ_c, λ_z, λ_u, β

Steps:

KL and recostr_loss → **1.** Initialize the VAE by minimizing L_{vae} on X, with c sampled from prior $p(c)$

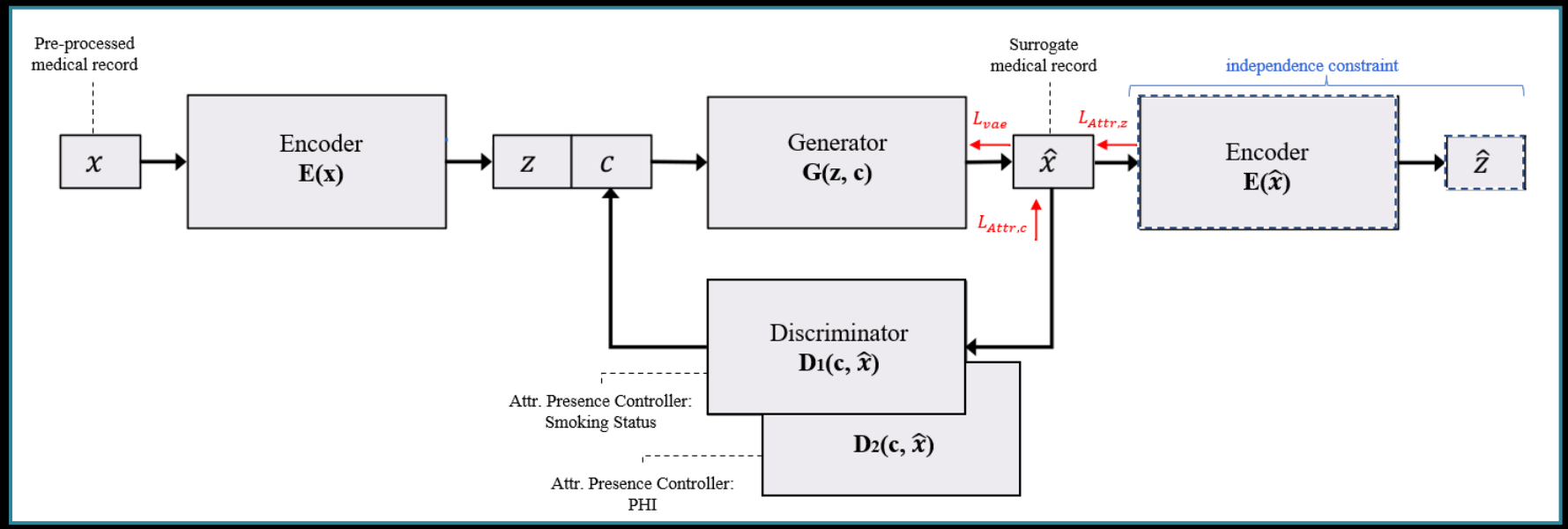
repeat

Using fake and real samples → **2.** Train the discriminator D_i by minimizing L_u

CTG → **3.** Train the generator G and the encoder E by minimizing L_G and L_{vae}

until convergence

Output: trained text generator G conditioned on disentangled representation (z, c)



Unrolled pipeline

DATA

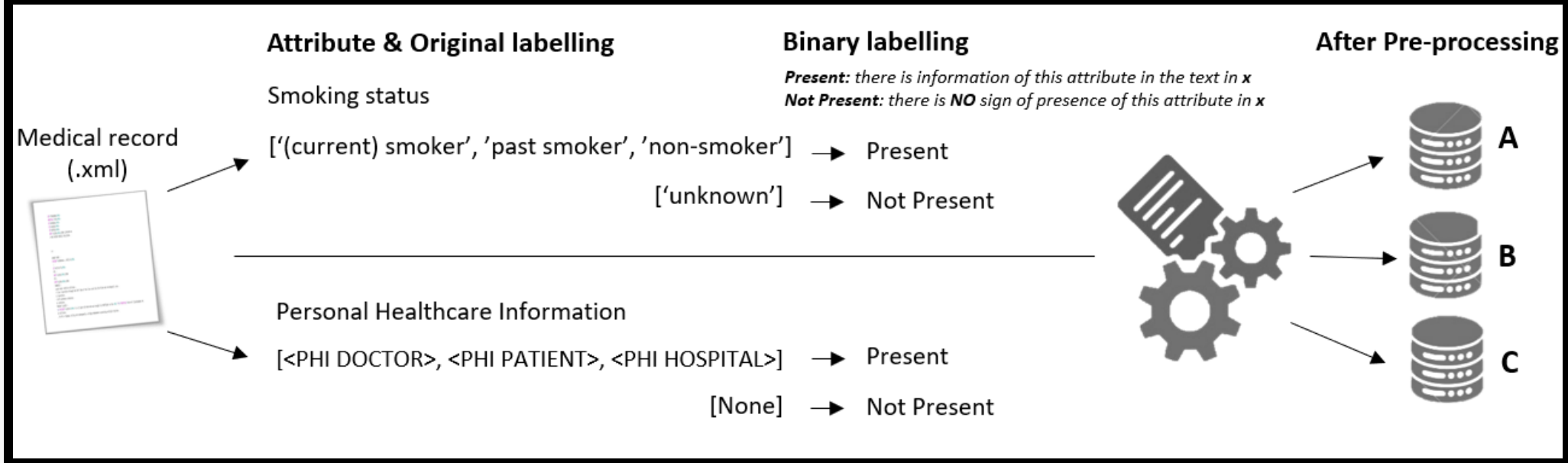
Automatic de-identification & smoking-status classification of Clinical Records

N2C2 2006

i2b2

669 de-identified records, annotated with **8** different categories of PHI.
502 records labelled with one of the **4** smoking-status categories.

.xml to .txt:
- Get PHI tags
- Extract headers



Dataset A: A.1.with headers and A.2 without (**669** records for training)

Dataset B: 1824 chunks o text from A.2 labelled for PHI presence, no headers

Dataset C: Without headers (**398** records for training)

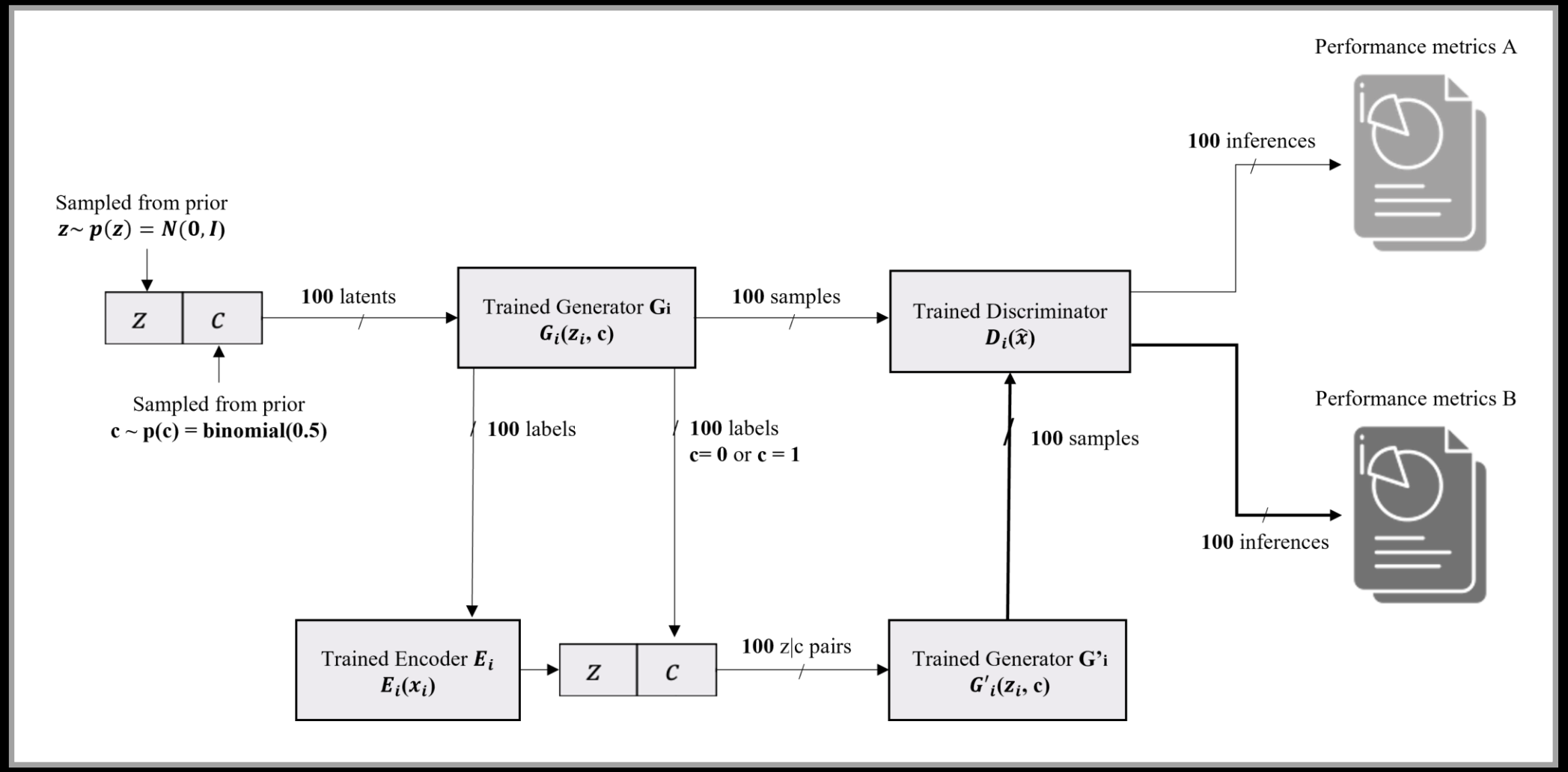
Experiment 1
Conditioning Smoking Status Presence

- Output Train Generator and Discriminator G1 and D1
- Trained with Dataset A (both initially) during the first phase (long text)
- Trained with Dataset C during the second phase (re-labelled for smoking status presence)

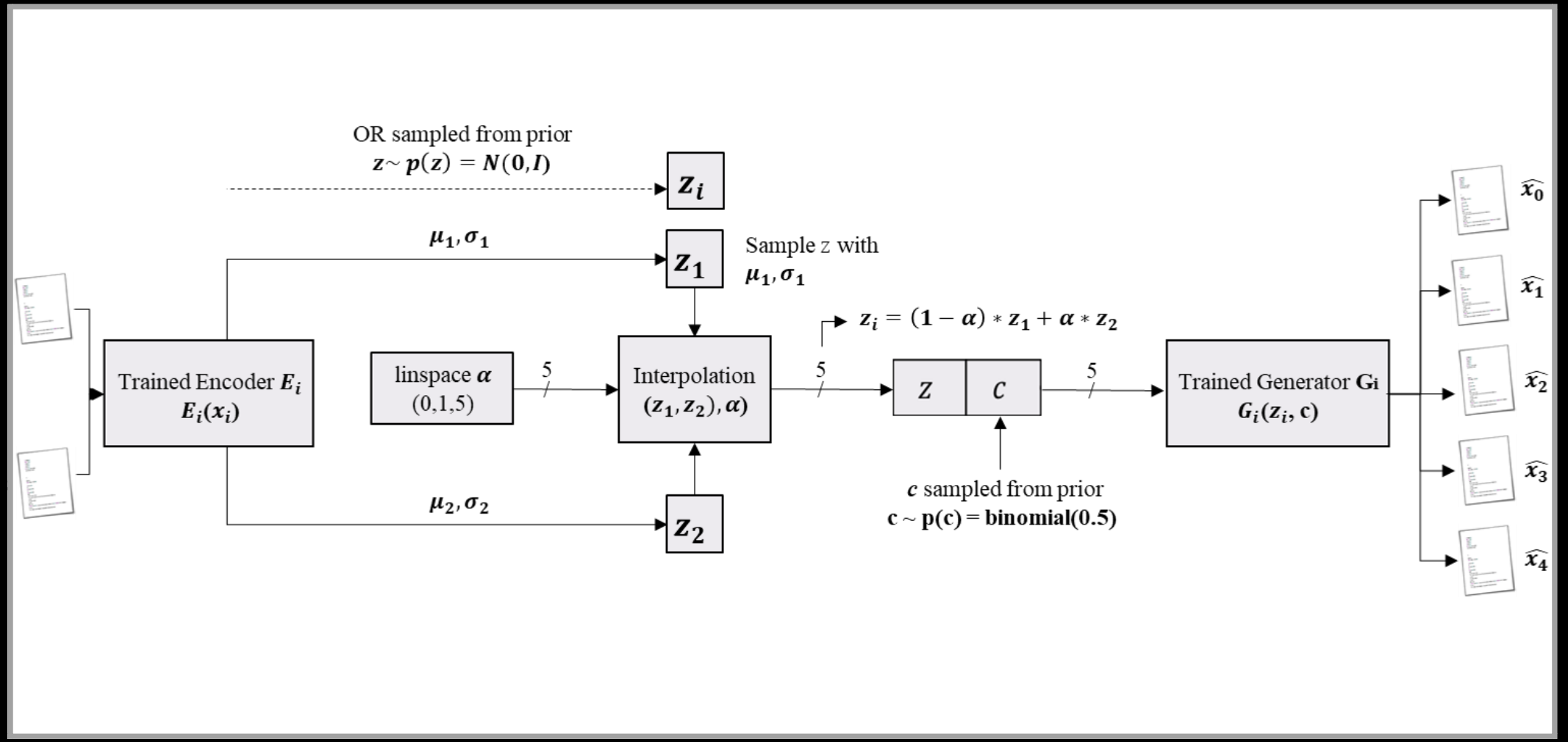
Experiment 2
Conditioning PHI Presence

- Output Train Generator and Discriminator G2 and D2
- Trained with Dataset A (A.2) during the first phase (long text)
- Trained with Dataset B during the second phase (labelled for PHI presence)

CONDITIONING EFFICACY



INTERPOLATION



RESULTS

metric \ target	D1: smoking-status presence	D2: PHI presence
Accuracy	0.90	0.93
Precision	0.87	0.87
Recall	0.87	0.86
Specificity	0.92	0.95
F1	0.81	0.86
Support	27/53	90/275

Table 4.1: Performance metrics for trained discriminators against test data)

metric \ target	D1_prior	D1_ctrl	D2_prior	D2_ctrl
Accuracy	0.45	0.75	0.47	0.47
Precision	0.50	0.87	0.50	- *(0/0)*
Recall	0.36	0.86	0.07	0.00
Specificity	0.55	0.91	0.45	1.0
F1	0.41	0.86	0.12	- *(0/0)*
Support	11/9		53/47	

Table 4.2: Conditioning mechanism efficacy. Improved performance metrics via conditioning.

RESULTS

REAL RECORD CONTAINING SMOKING STATUS INFORMATION

,812367409SH21952193

06/19/1991 12:00:00 AM

Discharge SummarySignedDIS

Admission Date :06/19/1991

Report Status :Signed

Discharge Date :

Independence Day

HISTORY OF PRESENT ILLNESS :

The patient is an 84-year-old woman with a history of rheumatoid arthritis .She is now status post three myocardial infarctions .She has had progressive deformity and rheumatoid arthritis of her right knee .She presented at this time for a right total knee replacement .

PAST MEDICAL HISTORY :

As above .Appendectomy .Cholecystectomy .Left total knee replacement in 1977 .Pepticulcer disease .MEDICATIONS :On admission included Inderal , 40 mgpo q.i.d. ; Aldomet , 250 mg po t.i.d. ; apresoline , 250 mg po t.i.d. ; Nitropaste , one - half inch q p.m. ; Zantac , 150 mg po q p.m. ; Lasix , 20 mg po q day ; allopurinol , 300 mg po b.i.d. and Clinoril , 25 mgpo b.i.d.HABITS :She does not smoke or drink .

ALLERGIES :

ECOTRIN .MINIPRES .TAGAMET .HALDOL .

PHYSICAL EXAMINATION :

On admission revealed an elderly woman in no acute distress .Temperature 97.6 .Pulse 80 .Respiratory rate 18 .Blood pressure 200/86 .Skin was without rashes or breakdown .Lungs were clear to auscultation and percussion .Heart revealed a Grade IV / VI systolic ejection murmur .Abdomen was soft , nontender , no masses .Extremities revealed skin was intact to both lower extremities .Right leg demonstrated flexion from 5 degrees to 135 degrees .Left leg demonstrated flexion from 0 to 130 degrees .She had tenderness in her right knee .She had crepitus in her right knee .Sensory and motor function was intact .

LABORATORY DATA :

On admission included x-rays which demonstrated severe degenerative disease of her right knee .She had a creatinine of 2.0 .BUN 59 .Hematocrit of 31.4 .EKG revealed left bundle branch block with no acute ischemic changes .Urinalysis demonstrated a small amount of blood but no evidence of infection .

HOSPITAL COURSE :

Rheumatoid arthritis :The patient underwent a right total knee replacement after Cardiology clearance .She tolerated the procedure well , however , in the immediate postoperative period she developed confusion and adelirium status state .She was evaluated by Neurology and followed carefully .All of her pain medications were discontinued and she was maintained with a sitter .Psychiatry Service evaluated the patient and she was scheduled for a head CT .Head CT demonstrated no evidence of any stroke or acute compromise but there appeared to be chronic atrophy .Further evaluation consisted of a Urology consult which followed for her mildly elevated creatinine .The patient was followed by the Cardiology Service postoperatively and demonstrated no evidence of myocardial infarction in the immediate postoperative period .Note that she cleared mentally spontaneously over approximately seven days postoperatively .She subsequently did well with physical therapy .She was cleared for discharge after achieving flexion beyond 100 degrees .She was able to ambulate up and down stairs with crutches .She had x-rays taken confirming good alignment of her prosthesis .Ultrasound ruled out evidence of a deep venous thrombosis .She was therefore discontinued on Coumadin .

DISCHARGE DIAGNOSES :

STATUS POST RIGHT TOTAL KNEE REPLACEMENT .CONFUSION .CHRONIC RENAL FAILURE .

DISPOSITION :

The patient was discharged to home in satisfactory condition .

MEDICATIONS :

On discharge included Zantac , 150 mg po q p.m. ; Lasix , 20 mg po q day ; allopurinol , 300mg po b.i.d. ; Clinoril , 325 mg po b.i.d. ; Nitropatch , one - half inch q p.m. ; apresoline , 250 mg po t.i.d. ; Aldomet , 250 mg po t.i.d. ; Inderal , 40 mg po q.i.d.IL799/1282RAMAG L. TROISQUARCKAYS , M.D. KK5D :07/11/91Batch :8122Report :H7634J2T :07/15/91Dictated By :HEAGLE [M.D.[report_end] ed and ct scan on ct scan at consult but showed no active"

Sample from the *Automatic Smoking Status Classification Challenge*
Patient's label: Non-smoker

discharge summary unsigned dis report status :

unsigned admission date :

02/04/92

discharge date :

#02/15/92

history of present illness :

the patient is a 68 year old female with a history of **squamous cell carcinoma** which is **squamous cell carcinoma from the tongue resection to the primary surgical resection**
The patient 's history on admission on admission , .the patient presented with a gastric collection .

history of present illness :

the patient is a seventy - year - old white female .the following history of left shows chest x - large 4 - 11 . same day exploration of a mass parathyroid adenoma , pelvic tube placement .the patient was admitted for anticoagulation for the motorvehicle with positive and three weeks of chest mass .on rectal presented in this operation of a rectal biopsy .the patient 's preoperative ultrasound was palpable in palpable thyroid bleeding .on june 11 , 1994 , a cold at secondary to be a flexible junction of the first pelvis .she was brought to a previous surgical greater than right vertex of **cisplatin 160** .an ultrasound showed a papanicolaou smear negative .her fluid blood and was notable for a previously normal forseveral months with greasy ovarian mass .the patient had the first postoperatively when she underwent gynecologic and she had an uneventful night of " .her previous first 2 .on postoperative day two the patient complained of abdominal distention .the lower extremity only and regular bowel movement .a chest x - ray but no other than 9 and distention in her right lower extremity and down in the third hospital day five and a regular bowel movement .a chest x - rayon her left chest pain has no esophageal laceration movement .she has had no edema with left non - tender .on surgery and this was brought back to the same side after may of 1994 , the left four . she was previously therefore startedon the morning of admission before the day of admission that she had an episode of bowel pressure of brought over the steroids and was maintained given with steroids .

past medical history :

significant for a single - removal . she had insulin dependent asthma , and had an normal risk for for glaucoma to manual family wishes well without report without any angina were no intubation , although no respiratory distress .she has lost respiratory .cranial nerves developed , consistent with bilateral upper body edema .her lungs were clear to auscultation and percussion bilaterally .it was a regular , with a right leg left coronary artery .she demonstrates severe dysfunction and anasarca ; 70 % , hematocrit .her complete calcium was 1.7 , creatinine 1.0 .the hematocrit .her coagulation .chest showed not to percussion and right heart bibasilar aldosterone , the liver , still showed an ef of nonspecific enlarged right upper ejection fraction to v4 and the v4 ii within normal limits .an electrocardiogram changes .she was ruled out for severe stenosis of an ef of 60 .an infectious process cardiovascular care other cardiovascular tests by cardiac study was then discontinued and which was then subsequently transferred to an infectious workup which was non - diagnostic .she ruled out for further cardiac risk history for her hypotension and ruled out) on anti - 93 percent after the previously had an apparent episode of hypotension and some sinus tachycardia and a heart heart murmur .the patient ruled out for left lower lobe lesion artery lesion on this lesion showed a slight right lesion was lesion on myocardial infarction .she had no hemodynamic , ventricular function and heart failure , fluid collections were increased that was slightly and arm kept , a motor vehicle contained , air protein , consistent with progressive edema .the patient remained afebrile and developed some mild acute respiratory hematocrit .she remained on triple flexion with chronic steroid required .as mentioned , which her left lower extremity with the left leg for cbc , which was followed by a knee and which she received lap coli hematocrit of clears to baseline post - no pneumothorax and the limits which were within normal limits .her wounds healed some well lower extremity and was placed on decadron postoperatively and she was also noted in her postoperative day , and developed her increased amount of the biopsy in the left lower extremity and showed kept npo for her electrolytes paresis .secondary diagnoses on discharge : the patient was a deep venous thrombosis is to be due to transfer to her discharge are increased .

condition on discharge :

good .

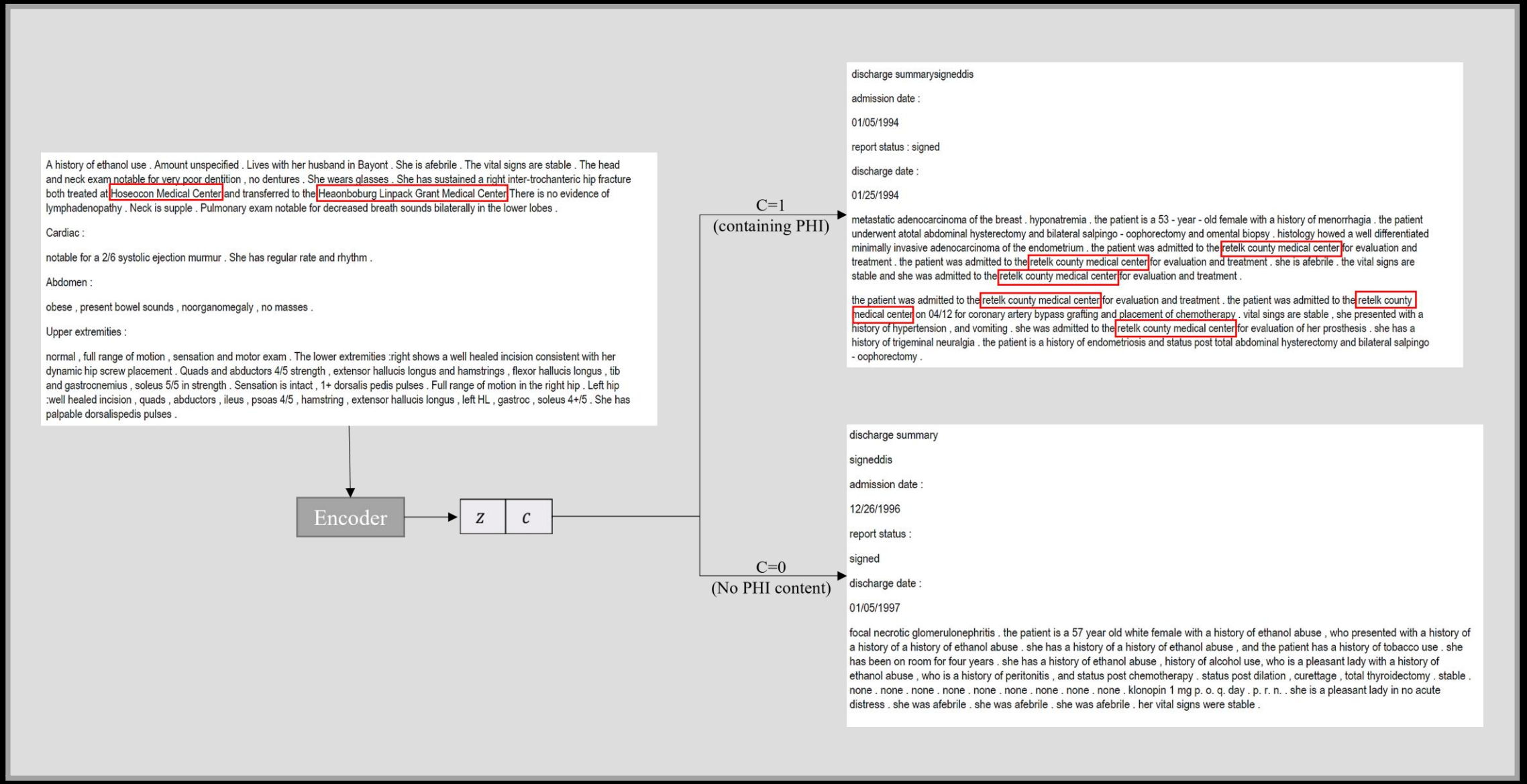
discharge diagnoses :

hypertension is a 5 - 5 - degree differential .left depression .left tube .left tube plantar strength / p .left tube :1 . right upper extremity used from her motor nodule in the left upper extremity used with a 75 pack

RESULTS

Sample from the *Automatic Smoking Status Classification Challenge*
Label: Smoking Information PRESENT

RESULTS



Conditioning on PHI Presence in the text.

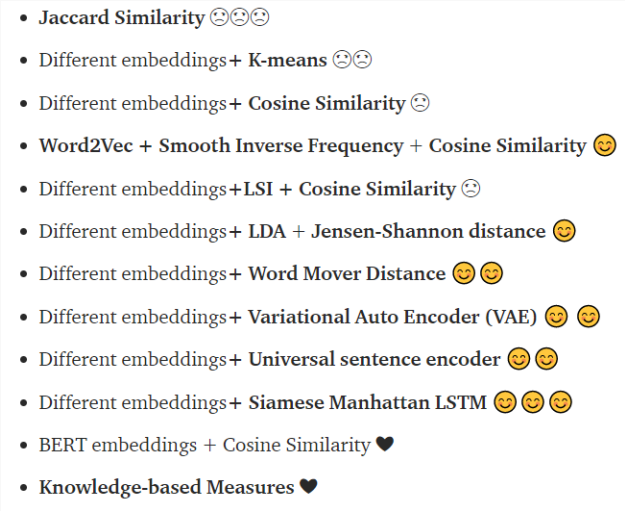
Thank you.

Remarks/Challenges

- Error Analysis and Implementation Complications
  - *Exploding gradients && Vanishing gradients in the same network*
  - *Next token search algorithm: beam search & when to stop.*
  - *objective definition, early_stopping, data pre-processing, feature engineering*

- Scalability to FL and “fully” ctg-vae.
  - *Discriminators for different attributes can be trained independently on separate labelled sets.*
  - *Method is able to effectively lift the word level knowledge to sentence level and generate convincing text.*

- Further Work
  - *Hyperparameter Tuning*
  - *More data*
  - *More features (separate PHI)*
  - *More dimensions (from binary to multilabel)*
  - *More discriminators in training*



Output

- Generative model based on VAE.
  - Generate novel medical records (sampled from random): data augmentation in Healthcare
  - Generate surrogate medical records x' very similar to real x

  Note: sampling text from a continuous space

- Discriminative models based on TextCNN:
  - Smoking-status-information-presence binary classifier: __% acc.
  - PHI-presence binary classifier: __% acc.

- Controlled text generation using CVAE-TextCNN:
  - Generate a novel medical record or a substitute medical record (based on x) controlling the following features.
    - Feature 1: if surrogate/novel medical record contains or not information regarding smoking-status/habits of the patient.
    - Feature 2: if surrogate/novel medical record contains or not protected healthcare information.

- Interpolation and Interpretability:
  - *This model is able to estimate an unknown value from a set of sample points with known values.*
  - *Sampling from a highly regular continuous space is useful to interpret why Discriminator ½ assigns a label to a specific real/fake sample (sampling from points very close to the poi)*

DATASETS

Automatic de-identification & smoking-status classification of Clinical Records

N2C2 2006

- **502** de-identified medical discharge records
- Annotated with **8** different categories of PHI
- Single label from **4** smoking-status categories

- **112.000** de-identified medical discharge records
- Each record with **7** ICD-9 codes on average

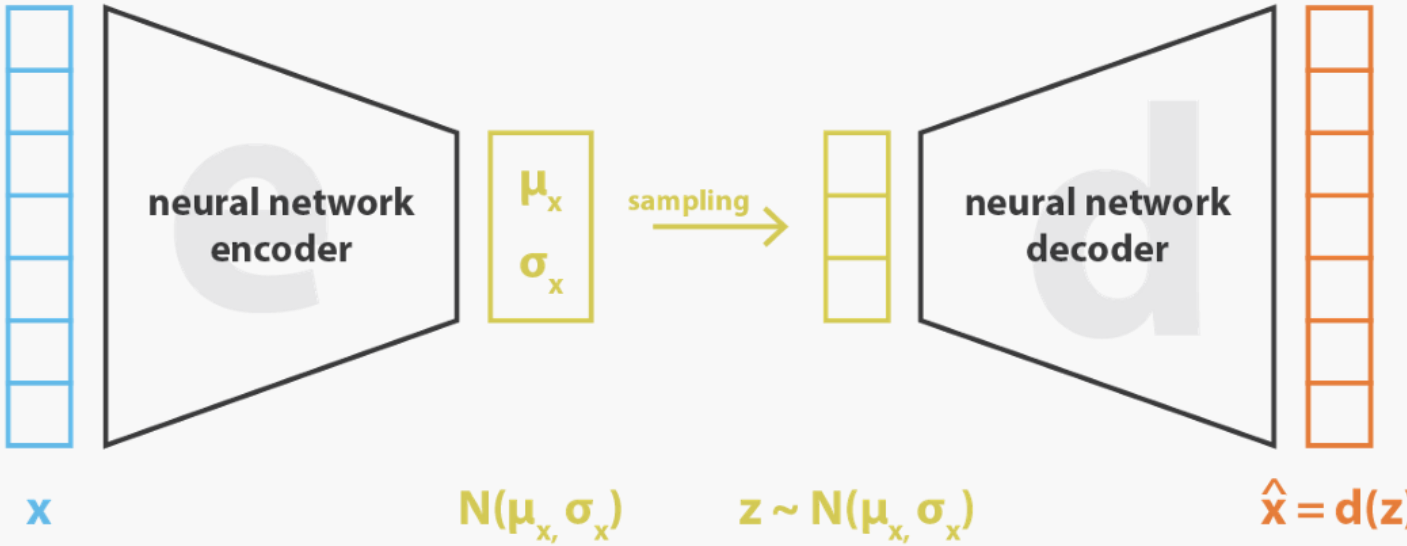
2006

	i2b2	MIMIC
Vocabulary size	46,803	69,525
Number of notes	1,304	1,635
Number of tokens	984,723	2,945,228
Number of PHIs	28,867	60,725
Number of PHI tokens	41,355	78,633

...

- 2014 i2b2: **800** records, **8** PHI categories
- 2016 CEGS N-GRID NLP: **1,000** records, 33,000 PHI instances

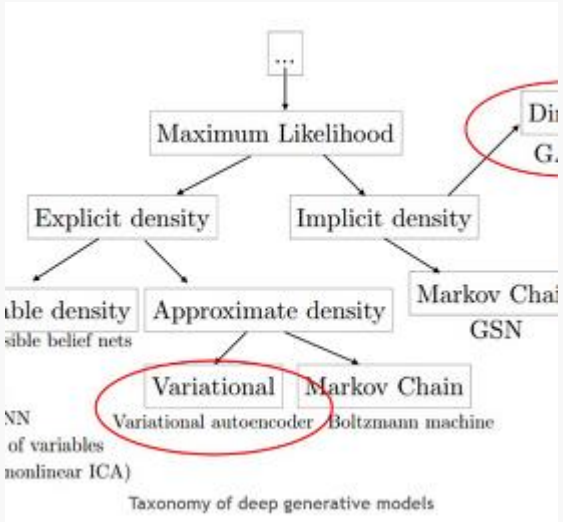
ARQUITECTURE (VAE)



$$\text{loss} = \|x - \hat{x}\|^2 + KL[N(\mu_x, \sigma_x), N(0, I)] = \|x - d(z)\|^2 + KL[N(\mu_x, \sigma_x), N(0, I)]$$

In variational autoencoders, the loss function is composed of a reconstruction term (that makes the encoding-decoding scheme efficient) and a regularisation term (that makes the latent space regular).

Technique	Advantages	Disadvantages
RNN	Natural sequence structure is very suitable for the task of sequence modeling	Cannot effectively capture the long-distance dependence between sentences
GAN	Unsupervised learning; Generating clearer and more realistic samples than other generative models	Instable training process; Not suitable for processing discrete data, such as text
Reinforcement learning	Similar to human learning manners; Combining with GAN can subtly solve the existing problems in GAN and generate realistic text	Quite complicated training process
VAE	Leveraging the latent vectors to increase the diversity of the generated text	The latent variable ensures that the desired content is generated, regardless of its quality
Transformer	The attention mechanism can efficiently capture the long-term context information; Fast parallel computing speed	Large amount of calculation and slow training speed



Taxonomy of deep generative models

Table 1. Comparision between two models

Criteria	VAE	DCGAN
Learning type	Semisupervised & unsupervised	Unsupervised
Architecture	Convolutional Autoencoder	Convolutional networks with some constraint
Gradient Update	SGD with update to reconstruction and KL loss	SGD update to both Generator and Discriminator
Optimizer	Adam	Adam
Objective	Inference by matching latent data distribution to original data distribution	Learn structural hierarchy of objects in Generator and Discriminator
Performance Metrics	Log-likelihood and error rate	Accuracy and error rate

DATA

i2b2

Dataset A.1: With headers (**669** records for training)

<RECORD ID="635">
<TEXT>
<PHI TYPE="ID">779810048</PHI>
<PHI TYPE="HOSPITAL">FIH</PHI>
<PHI TYPE="ID">8956861</PHI>
<PHI TYPE="ID">641681</PHI>
<PHI TYPE="ID">027815</PHI>
<PHI TYPE="DATE">12/02</PHI>/1998 12:00:00 AM
(1) STATUS POST MOTOR VEHICLE COLLISION .
(2) GRADE
Unsigned
DIS
Report Status :
Unsigned
DISCHARGE SUMMARY NAME :
<PHI TYPE="PATIENT">CHIRDVAIA , RITOC M</PHI>
UNIT NUMBER :
<PHI TYPE="ID">427-83-75</PHI>
ADMISSION DATE :
<PHI TYPE="DATE">12/02</PHI>/1998
DISCHARGE DATE :
<PHI TYPE="DATE">12/02</PHI>/1998
PRINCIPAL DIAGNOSIS :
(1) Status post motor vehicle collision .
(2) Grade 4 liver laceration through the left lobe of the liver and into the hilum and retrohepatic cava
(3) Splenic laceration .
(4) Severe left pulmonary contusion .
(5) Cardiac contusion .
HISTORY OF PRESENT ILLNESS :
Mr. <PHI TYPE="PATIENT">Cuchsli</PHI> is a 17-year-old male who was brought via Medflight to the <PHI TYPE=
motor vehicle collision .
Per report , he hit a highway railing and subsequently a bridge embankment sustaining multiple injuries .
Per report , there was extensive damage to the vehicle and he required a very prolonged extrication time .
His accident occurred , per report , at 6:45 in the morning , and he was brought to the <PHI TYPE="HOSPITAL

Sample from the de-identification challenge

HISTORY OF PRESENT ILLNESS :

The patient is an 84-year-old woman with a history of rheumatoid arthritis .She is now status post three myocardial infarct
presented at this time for a right total knee replacement .

PAST MEDICAL HISTORY :

As above .Appendectomy .Cholecystectomy .Left total knee replacement in 1977 .Pepticulcer disease .MEDICATIONS :C
250 mg po t.i.d. ; Nitropaste , one - half inch q p.m. ; Zantac , 150 mg po q p.m. ; Lasix , 20 mg po q day ; allopurinol , 300

ALLERGIES :

ECOTRIN .MINIPRES .TAGAMET .HALDOL .

PHYSICAL EXAMINATION :

On admission revealed an elderly woman in no acute distress .Temperature 97.6 .Pulse 80 .Respiratory rate 18 .Blood pr
and percussion .Heart revealed a Grade IV / VI systolic ejection murmur .Abdomen was soft , nontender , no masses .Ext
flexion from 5 degrees to 135 degrees .Left leg demonstrated flexion from 0 to 130 degrees .She had tenderness in her ri

LABORATORY DATA :

On admission included x-rays which demonstrated severe degenerative disease of her right knee .She had a creatinine o
ischemic changes .Urinalysis demonstrated a small amount of blood but no evidence of infection .

HOSPITAL COURSE :

Rheumatoid arthritis :The patient underwent a right total knee replacement after Cardiology clearance .She tolerated the p
confusion and adelirium status state .She was evaluated by Neurology and followed carefully .All of her pain medications
the patient and she was scheduled for a head CT .Head CT demonstrated no evidence of any stroke or acute compromis
consult which followed for her mildly elevated creatinine .The patient was followed by the Cardiology Service postoperativ
postoperative period .Note that she cleared mentally spontaneously over approximately seven days postoperatively .She
achieving flexion beyond 100 degrees .She was able to ambulate up and down stairs with crutches .She had x-rays taken
venous thrombosis .She was therefore discontinued on Coumadin .

Without PHI tags,

Dataset B: 1824 shuffled chunks of text from **A.2** labelled for PHI presence; no headers

A history of ethanol use . Amount unspecified . Lives with her husband in Bayont . She is afebrile . Th
and neck exam notable for very poor dentition , no dentures . She wears glasses . She has sustained
both treated at Hoseocon Medical Center and transferred to the Heaonboburg Linpack Grant Medical
lymphadenopathy . Neck is supple . Pulmonary exam notable for decreased breath sounds bilaterally

Cardiac :

notable for a 2/6 systolic ejection murmur . She has regular rate and rhythm .

Abdomen :

obese , present bowel sounds , noorganomegaly , no masses .

Upper extremities :

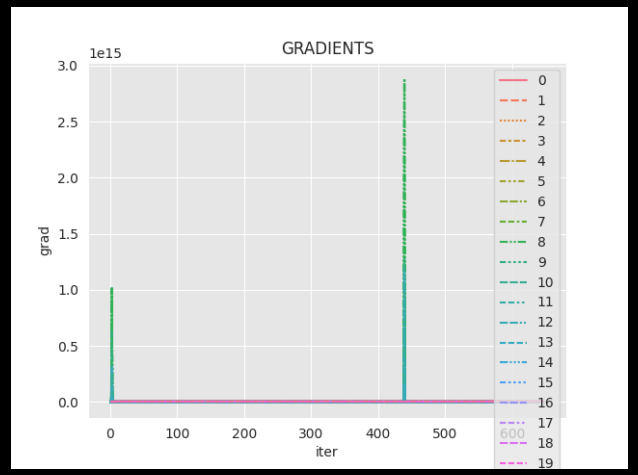
normal , full range of motion , sensation and motor exam . The lower extremities :right shows a well h
dynamic hip screw placement . Quads and abductors 4/5 strength , extensor hallucis longus and ham
and gastrocnemius , soleus 5/5 in strength . Sensation is intact , 1+ dorsalis pedis pulses . Full range
:well healed incision , quads , abductors , ileus , psoas 4/5 , hamstring , extensor hallucis longus , left
palpable dorsalispedis pulses .

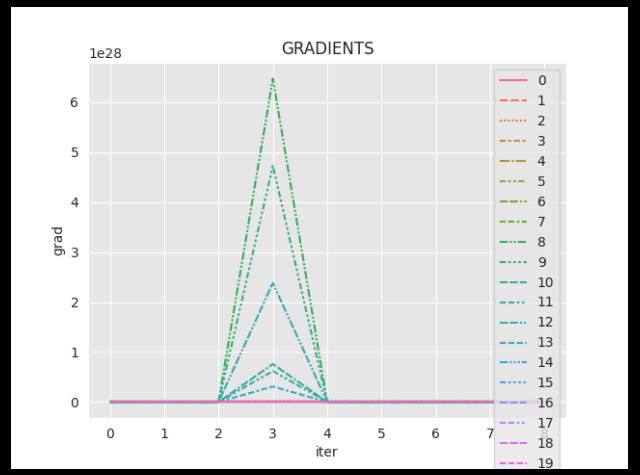
Chunked and labelled

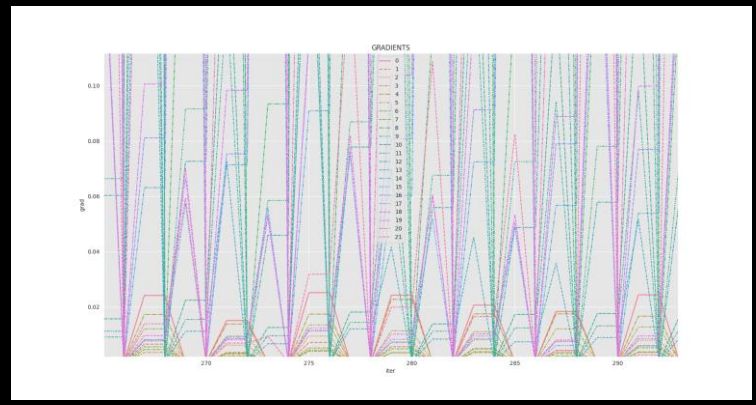
Monitoring Gradients

- *N* derivatives will be multiplied together in a network this big. If the derivatives are large, the gradient will increase exponentially as we propagate down the model until they eventually **explode**.

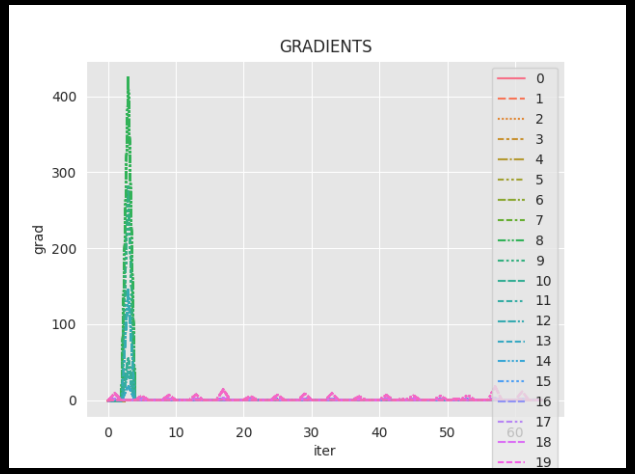
- Alternatively, if the derivatives are small then the gradient will decrease exponentially as we propagate through the model until it eventually **vanishes**.

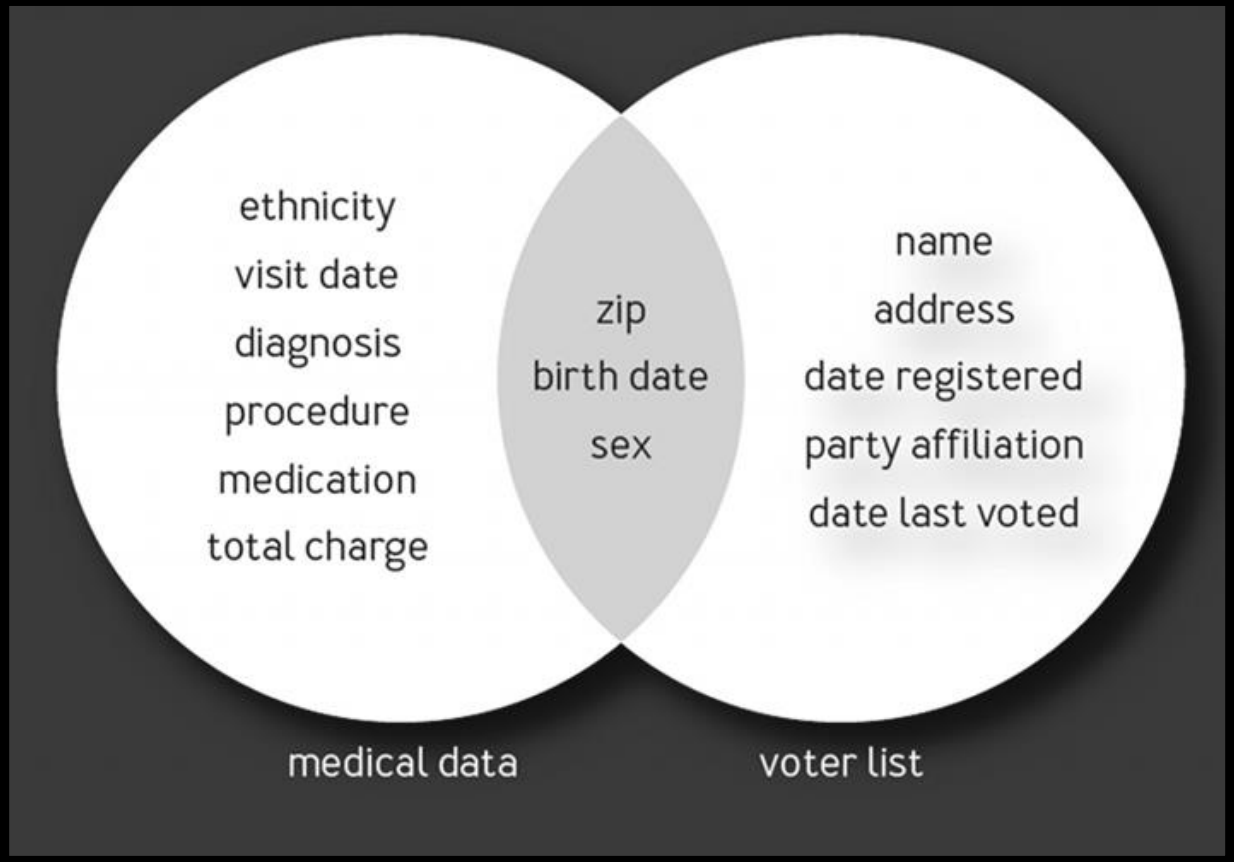






After gradient clippling, low learning rate (lr-4), and small batch size (16) →




ethnicity
visit date
diagnosis
procedure
medication
total charge
zip
birth date
sex
name
address
date registered
party affiliation
date last voted
medical data
voter list