
Finetuning Transformer Models

Per Egil Kummervold



Norway's petabyte plan: Store everything ever published in a 1,000-year archive

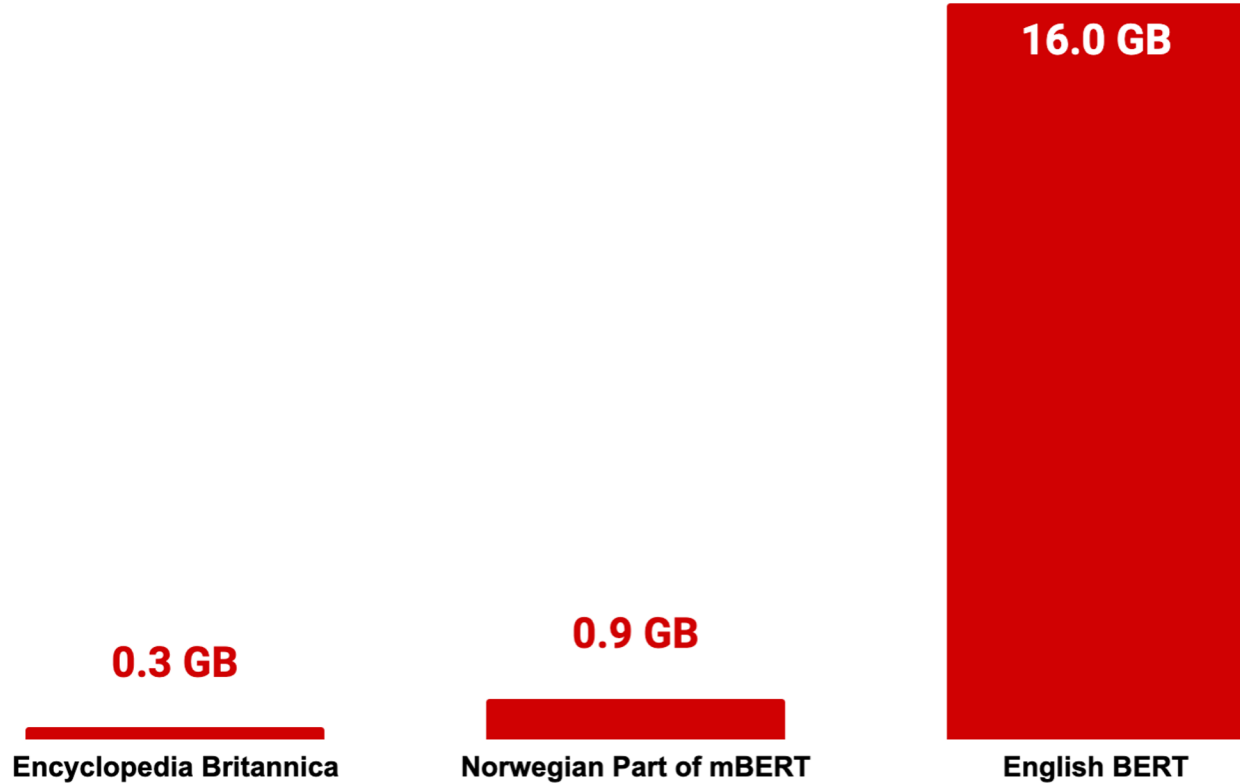
From ancient manuscripts to movies, the National Library of Norway wants to put it all online for the public.

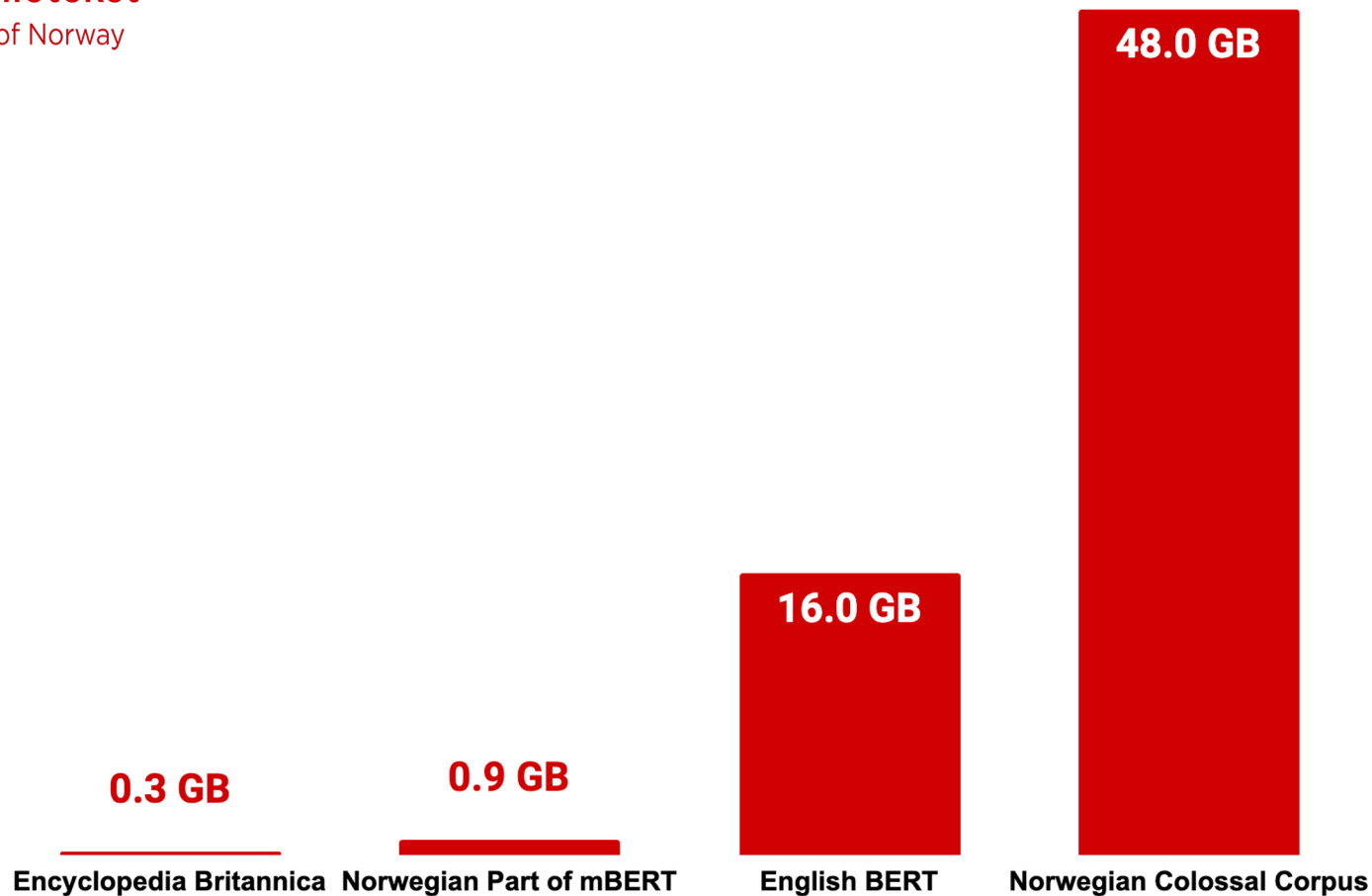


Started 2008

Status August 2022

88%	724.000	books
75%	5.180.000	newspapers
4%	2.000.000	magazines
50%	3.391.000	photos
58%	1.083.000	radio hours





NB Language Models

Encoder models

- Norwegian BERT base - nb-bert-base
- Norwegian BERT large - nb-bert-large
- Norwegian Sentence Transformer - nb-sbert-base

Encoder-decoder models

- NorthT5 small/base/large/XL/XXL
- Coming: nb-T5 models

Decoder model

- Norwegian GPT-J (6B)

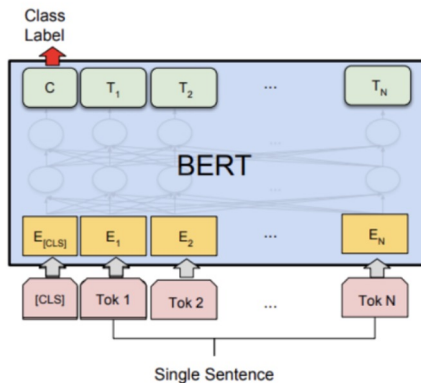
+ **Speech models**

Pretraining vs finetuning

- Pretraining is (typically) done unsupervised on masked text
- Finetuning is (usually) task specific
- Finetuning is (usually) done on moderate hardware
- Finetuning is (usually) done on moderate (annotated) dataset
- Domain Specific Pre-training
- Both pretraining and finetuning varies between model variants

Fine-tuning an Encoder-only Model

- Classification tasks
- Add a fully connected layer before the class labels, and train the entire model

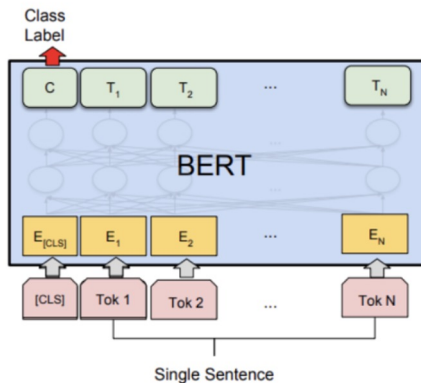


(b) Single Sentence Classification Tasks:
SST-2, CoLA

Review	Sentiment
This movie was amazing, I loved every minute of it.	Positive
The acting was terrible, I couldn't stand it.	Negative
I found this movie to be mediocre at best.	Neutral
This film had great special effects, but a weak plot.	Positive
The dialogue was witty and clever, I really enjoyed it.	Positive

Fine-tuning an Encoder-only Model

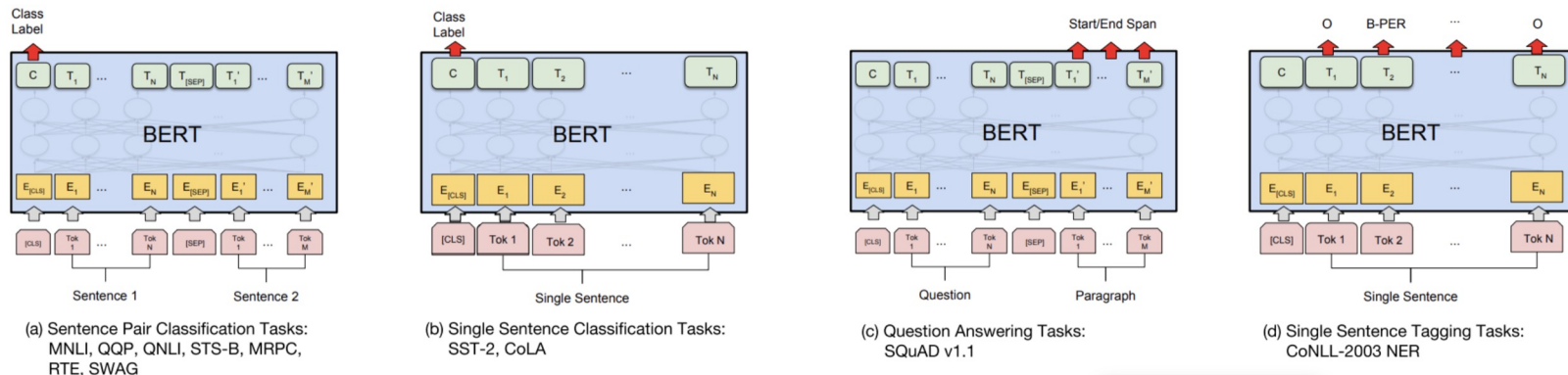
- Classification tasks
- Add a fully connected layer before the class labels, and train the entire model



(b) Single Sentence Classification Tasks:
SST-2, CoLA

Review	Sentiment
This movie was amazing, I loved every minute of it.	2
The acting was terrible, I couldn't stand it.	0
I found this movie to be mediocre at best.	1
This film had great special effects, but a weak plot.	2
The dialogue was witty and clever, I really enjoyed it.	2

Fine-tuning an Encoder-only Model



- Hard to use on more complex NLP tasks
- Other methods can give superior results

Fine-tuning an Encoder-only Model

DON'T

Fine-tuning an Encoder-only Model

DON'T

except research and direct use cases

Fine-tuning a Decoder-only Model

- Only the big models (+100B parameters) are really good
- Finetuning a 100B model is not feasible for ordinary mortals
- Training a small model in front (prompt tuning)
- Training a separate model to evaluate the output (RLHF)
- Challenges
 - Stochastic models
 - Long sequences might be unpredictable

Fine-tuning a Decoder-only Model

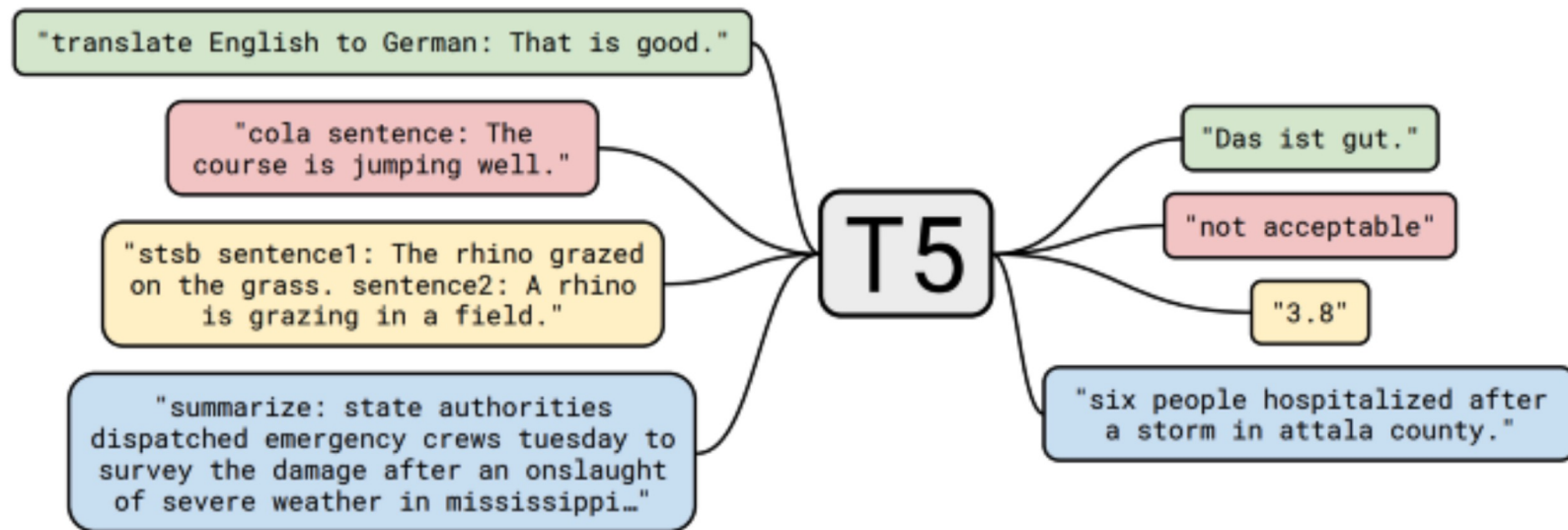
\$\$\$



T5 (Text-To-Text Transfer Transformer)

- Available in:
 - small - 60M
 - base - 220M
 - large - 770M
 - XL - 3B
 - XXL - 11B
- Smaller models can be fine-tuned in Colab

Finetuning a Encoder-Decoder Model



Finetuning an Encoder-Decoder Model

Review	Sentiment
This movie was amazing, I loved every minute of it.	Positive
The acting was terrible, I couldn't stand it.	Negative
I found this movie to be mediocre at best.	Neutral
This film had great special effects, but a weak plot.	Positive
The dialogue was witty and clever, I really enjoyed it.	Positive

Finetuning a Encoder-Decoder Model

Source Text (Norwegian)	Target Text (English)
Hei, hvordan har du det i dag?	Hi, how are you doing today?
Jeg skal dra på ferie til Spania neste uke.	I'm going on vacation to Spain next week.
Kan du hjelpe meg med å oversette denne setningen?	Can you help me translate this sentence?
Jeg er veldig sulten, la oss spise middag nå.	I'm very hungry, let's eat dinner now.
Takk for hjelpen, jeg setter stor pris på det.	Thank you for your help, I really appreciate it.

Finetuning on translation

Instruction tuning?

Source Text (Norwegian)	Target Text (English)
translate Norwegian to English: Hei, hvordan har du det i dag?	Hi, how are you doing today?
translate Norwegian to English: Jeg skal dra på ferie til Spania neste uke.	I'm going on vacation to Spain next week.
translate Norwegian to English: Kan du hjelpe meg med å oversette denne setningen?	Can you help me translate this sentence?
translate Norwegian to English: Jeg er veldig sulten, la oss spise middag nå.	I'm very hungry, let's eat dinner now.
translate Norwegian to English: Takk for hjelpen, jeg setter stor pris på det.	Thank you for your help, I really appreciate it.

Finetuning on summarization

Source Text	Target Text
summarize text: The United States is a diverse country with a rich cultural heritage and complex political system. It is known for its history of immigration and commitment to democratic values...	The US is diverse with rich cultural heritage and complex political system.
summarize text: The COVID-19 pandemic has had significant global impact, leading to widespread illness and economic disruption...	The COVID-19 pandemic has had significant global impact.
summarize text: Machine learning is a rapidly growing field that uses algorithms to enable computers to learn from data, with applications in natural language processing, computer vision, and speech recognition...	Machine learning is a rapidly growing field with many applications.
summarize text: The history of the Roman Empire has influenced modern Western civilization in areas such as law, government, and architecture...	The Roman Empire has influenced modern Western civilization.

Finetuning on language correction

Sentence	Corrected Sentence
correct errors in text: I likes to play video games.	I like to play video games.
correct errors in text: The car's break wasn't working.	The car's brake wasn't working.
correct errors in text: He study economics in college.	He studies economics in college.
correct errors in text: She dranked coffee every morning.	She drank coffee every morning.
correct errors in text: Their always late for meetings.	They're always late for meetings.

Fine-tuning without a fine-tuning dataset

source_text	target_text
The quick brown fox jumps over the lazy dog.	The quick brown fox jumps over the lazy dog.
She woke up late, feeling groggy from lack of sleep.	She woke up late, feeling groggy from lack of sleep.
The smell of freshly-baked bread wafted through the air.	The smell of freshly-baked bread wafted through the air.
He nervously twirled his hair around his finger, waiting for his turn.	He nervously twirled his hair around his finger, waiting for his turn.
The sun sets over the mountains in a brilliant display of colors.	The sun sets over the mountains in a brilliant display of colors.

... remove comma

source_text	target_text
The quick brown fox jumps over the lazy dog.	The quick brown fox jumps over the lazy dog.
She woke up late feeling groggy from lack of sleep.	She woke up late, feeling groggy from lack of sleep.
The smell of freshly-baked bread wafted through the air.	The smell of freshly-baked bread wafted through the air.
He nervously twirled his hair around his finger waiting for his turn.	He nervously twirled his hair around his finger, waiting for his turn.
The sun sets over the mountains in a brilliant display of colors.	The sun sets over the mountains in a brilliant display of colors.

Fine-tuning Named Entities

Sentence	NER Tags
identify entities in text: Tom works for a software company in San Francisco.	PER O O O O O LOC LOC.
identify entities in text: The Great Wall of China is one of the Seven Wonders of the World.	O O LOC LOC O O O O O.
identify entities in text: Jane is a doctor and lives in Toronto with her family.	PER O O O O O LOC O O.
identify entities in text: The Mona Lisa was painted by Leonardo da Vinci in the 16th century.	O ART O O PER PER O O O TIME.
identify entities in text: Apple is a technology company based in Cupertino, California.	ORG O O O O LOC LOC.

Fine-tuning Named Entities

Sentence	Marked-up Sentence
identify persons in text: Tom works for a software company in San Francisco.	<PER>Tom</PER> works for a software company in San Francisco.
identify persons in text: The Great Wall of China is one of the Seven Wonders of the World.	The Great Wall of China is one of the Seven Wonders of the World.
identify persons in text: Jane is a doctor and lives in Toronto with her family.	<PER>Jane</PER> is a doctor and lives in Toronto with her family.
identify persons in text: The Mona Lisa was painted by Leonardo da Vinci in the 16th century.	The Mona Lisa was painted by <PER>Leonardo da Vinci</PER> in the 16th century.
identify persons in text: Apple is a technology company based in Cupertino, California.	Apple is a technology company based in Cupertino, California.

Fine-tuning Named Entities

source_text	target_text
identify persons in text: Tom works for a software company in San Francisco.	# works for a software company in San Francisco.
identify persons in text: The Great Wall of China is one of the Seven Wonders of the World.	The Great Wall of China is one of the Seven Wonders of the World.
identify persons in text: Jane is a doctor and lives in Toronto with her family.	# is a doctor and lives in Toronto with her family.
identify persons in text: The Mona Lisa was painted by Leonardo da Vinci in the 16th century.	The Mona Lisa was painted by # in the 16th century.
identify persons in text: Apple is a technology company based in Cupertino, California.	Apple is a technology company based in Cupertino, California.

Fine-tuning a Encoder-Decoder Model

source_text	target_text
Tom jobber for et programvareselskap i San Francisco.	# jobber for et programvareselskap i San Francisco.
Kinesiske muren er en av verdens syv underverker.	Kinesiske muren er en av verdens syv underverker.
Jane er en lege og bor i Toronto med familien sin.	# er en lege og bor i Toronto med familien sin.
Mona Lisa ble malt av # på 1500-tallet.	Mona Lisa ble malt av # på 1500-tallet.
Apple er et teknologiselskap med hovedkontor i Cupertino, California.	Apple er et teknologiselskap med hovedkontor i Cupertino, California.

might look easy but....

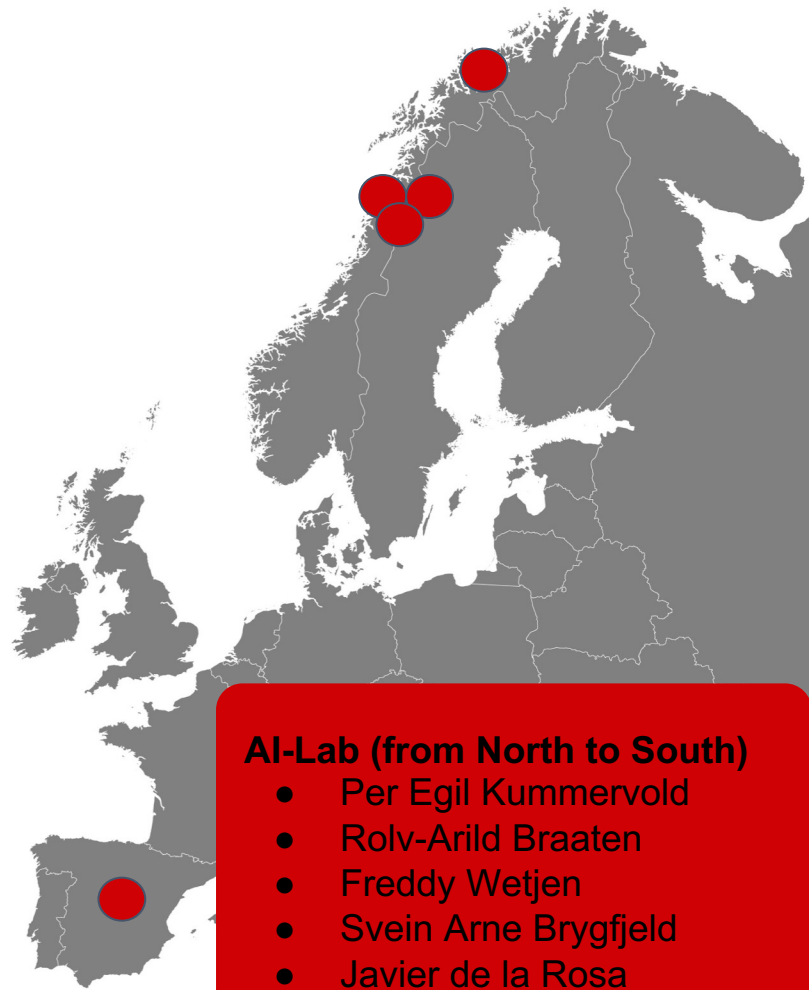
Named Entity Recognition

Per i dag har Per og Dag startet firmaet Per og Dag

Named Entity Recognition

Per i dag har Per og Dag startet firmaet Per og Dag

<https://ai.nb.no>



AI-Lab (from North to South)

- Per Egil Kummervold
- Rolv-Arild Braaten
- Freddy Wetjen
- Svein Arne Bryggjeld
- Javier de la Rosa

```
✓ 1 # Install simple-T5
2 !pip install simplet5

✓ 1 import pandas as pd
2
3 # Set local output directory
4 outputdir = "/content/outputs"
5
6 # Load the data
7 train_df = pd.read_csv("https://storage.googleapis.com/notram-public/finetune_datasets/norne_simple/train2.tsv",sep="\t")
8 test_df = pd.read_csv("https://storage.googleapis.com/notram-public/finetune_datasets/norne_simple/test.tsv",sep="\t")
9
10 train_df
```

1 to 25 of 15696 entries

Index	source_text	target_text
0	Lam og piggyvar på bryllupsmenyen.	Lam og piggyvar på bryllupsmenyen.
1	Kamskjell, piggyvar og lammefillet sto på menyen under den kongelige gallamiddagen.	Kamskjell, piggyvar og lammefillet sto på menyen under den kongelige gallamiddagen.
2	Og til dessert: Parfait à la Mette-Marit.	Og til dessert: Parfait à la Mette-Marit.
3	Forretten lyder navnet "Coquilles St. Jacques Prince de Norvege", som er grillt kamskjell på norsk spekeskinke - med trøffelhonningvinaigrette, ruccolasalat og ristede gresskarkjerner.	Forretten lyder navnet " Coquilles St. Jacques Prince de Norvege ", som er grillt kamskjell på norsk spekeskinke - med trøffelhonningvinaigrette, ruccolasalat og ristede gresskarkjerner.
4	Deretter fulgte "Turbot au Four - purée de chou-fleur".	Deretter fulgte " Turbot au Four - purée de chou-fleur ".
5	Det er ovnsbakt piggyvar dandert med ingefærmarinert klasetommatbiter, blomkålpuré drysset med finhakket gresslek, og fransk smørsaus med appelsin og fenikelsmak.	Det er ovnsbakt piggyvar dandert med ingefærmarinert klasetommatbiter, blomkålpuré drysset med finhakket gresslek, og fransk smørsaus med appelsin og fenikelsmak.
6	Hovedretten er "Filet d'agneau - girolles à la creme":	Hovedretten er " Filet d'agneau - girolles à la creme ":
7	Ytrefilet av norsk lam, kremet kantareller i nepe, bakt minisquash, leltrakte rødbeter, rissolepoteter og rosmarinsjy.	Ytrefilet av norsk lam, kremet kantareller i nepe, bakt minisquash, leltrakte rødbeter, rissolepoteter og rosmarinsjy.
8	Og desserten, "Parfait à la Mette-Marit - coulis de muron arctique" består av Snøfrisk og yoghurtisparfait med friske norske bringebær, blåbær og bjørnebær, samt multemousse på multecoulis.	Og desserten, " Parfait à la Mette-Marit - coulis de muron arctique " består av Snøfrisk og yoghurtisparfait med friske norske bringebær, blåbær og bjørnebær, samt multemousse på multecoulis.
9	Benket til bryllupsmiddag	Benket til bryllupsmiddag
10	Litt forsinket, klokken 20.25, satte bryllupsgjestene seg til bords for å nyte gallamiddagen på Slottet.	Litt forsinket, klokken 20.25, satte bryllupsgjestene seg til bords for å nyte gallamiddagen på Slottet.
11	I galla og utøkte festantrekk kom de til bryllupsmiddagen, kongelige, presidenter, familiemedlemmer, venner og representanter for regjering, storting og viktige norske institusjoner.	I galla og utøkte festantrekk kom de til bryllupsmiddagen, kongelige, presidenter, familiemedlemmer, venner og representanter for regjering, storting og viktige norske institusjoner.
12	De ble ønsket velkommen av dronning Sonja, kong Harald, kronprins Haakon og kronprinsesse Mette-Marit.	De ble ønsket velkommen av dronning #, kong #, kronprins # og kronprinsesse #.
13	Det kongelige bryllup er i høy grad også en privat begivenhet.	Det kongelige bryllup er i høy grad også en privat begivenhet.
14	Det viste prosenjonen til gallamiddagen på Slottet og plasseringen ved hovedbordet.	Det viste prosenjonen til gallamiddagen på Slottet og plasseringen ved hovedbordet.
15	Først i prosenjonen gikk kong Harald og brudens mor, Marit Tjessem.	Først i prosenjonen gikk kong # og brudens mor, ##.
16	Bak fulgte brudens far, Sven O. Heib, med dronning Sonja ved armen.	Bak fulgte brudens far, ##, med dronning # ved armen.
17	Først i tredje rekke kom svensk kongen, Carl Gustaf, og dronning Margrethe av Danmark.	Først i tredje rekke kom svensk kongen, ##, og dronning ###.
18	Det er ikke hverdagskost at monarkene i våre to naboland blir rangert bak vanlige norske statsborgere.	Det er ikke hverdagskost at monarkene i våre to naboland blir rangert bak vanlige norske statsborgere.
19	Men så er ikke bryllupet mellom Haakon og Mette-Marit hverdagskost det heller.	Men så er ikke bryllupet mellom # og # hverdagskost det heller.
20	Presidenter og prinser	Presidenter og prinser
21	Bak våre nabolands statsoverhoder fulgte Island's president Ólafur Ragnar Grímsson med dronning Silvia av Sverige.	Bak våre nabolands statsoverhoder fulgte Island's president ### med dronning ##.
22	Deretter storthertug Henri av Luxembourg med Finlands kvinnelige president, Tarja Halonen, så prins Albert av Monaco med den spanske dronning Sofia og nest sist i prosenjonen den nederlandske kronprinsen, Willem-Alexander med sin forlovede Maxima Zorreguieta.	Deretter storthertug ## med Finlands kvinnelige president, ##, så prins ### med den spanske dronning # og nest sist i prosenjonen den nederlandske kronprinsen, # med sin forlovede ##.
23	Forlovede par skiller ikke i en kongelig prosenjon, ei heller ved det kongelige taffel.	Forlovede par skiller ikke i en kongelig prosenjon, ei heller ved det kongelige taffel.
24	Prosenjonen ble avsluttet av den 90 år gamle prins Carl Bernadotte som førte brudgommens søster, Märtha Louise, til bords.	Prosenjonen ble avsluttet av den 90 år gamle prins ## som førte brudgommens søster, #, til bords.

Show per page

1 2 10 100 600 620 628

Like what you see? Visit the [data table notebook](#) to learn more about interactive tables.

✓
24m



```
1  from simplet5 import SimpleT5
2
3  # Load a pretrained model
4  model = SimpleT5()
5  model.from_pretrained(model_type="mt5", model_name="north/t5_base_NCC")
6
7  # Finetune the model
8  model.train(train_df=train_df[:,
9              eval_df=test_df[:,
10             source_max_token_len=64,
11             target_max_token_len=64,
12             batch_size=32,
13             max_epochs=5,
14             #early_stopping_patience_epochs = 0,
15             outputdir=outputdir,
16             use_gpu=True)
```



```
1 # Load the finetuned model for prediction
2 # Find the folder where the model is stored, and update the path below
3 model.load_model("mt5", "/content/output/model-test", use_gpu=True)
4
5 text_to_clean="Akkurat nå fullførte Per Egil denne presentasjonen."
6 model.predict(text_to_clean)
```