ESSIR
The 14th European Summer School on Information Retrieval

MSCA ITN/ETN No. 860721

DoSSIER
Domain Specific Systems for Information Extraction and Retrieval

Neural reranking

Georgios Peikos, Wojciech Kusa, Annisa Maulida Ningtyas, Oscar E. Mendoza

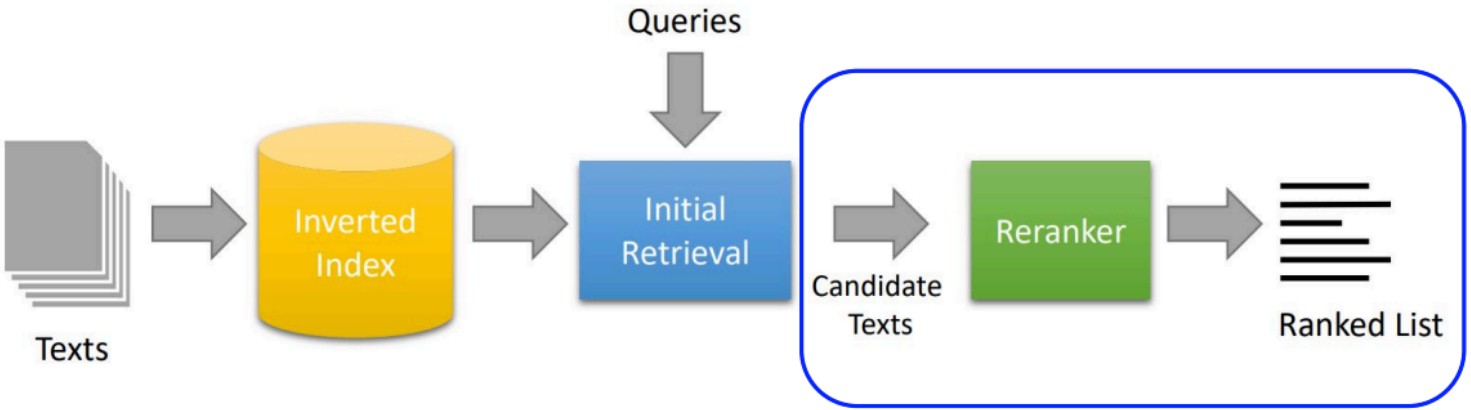
Agenda

1. Neural retrieval recap
2. Challenges in medical neural retrieval
3. Neural retrieval for clinical trials
4. Introduction to the TCRR model
5. Performance tracking
6. Hands-on #1: model training
7. Hands-on #2: inference

Agenda

1. Neural retrieval recap
2. Challenges in medical neural retrieval
3. Neural retrieval for clinical trials
4. Introduction to the TCRR model
5. Performance tracking
6. Hands-on #1: model training
7. Hands-on #2: inference

Neural Retrieval



DoSSIER MSCA ITN/ETN No. 860721

Three approaches to neural IR

1. Cross-encoder
2. Bi-encoder
3. Learned sparse retrieval

Agenda

1. Neural retrieval recap
2. Challenges in medical neural retrieval
3. Neural retrieval for clinical trials
4. Introduction to the TCRR model
5. Performance tracking
6. Hands-on #1: model training
7. Hands-on #2: inference

Technical Challenges

- Data Quality
- Interpretability
- Scalability
- Integration

Ethical and Regulatory Challenges

- Patient Privacy
- Bias and Fairness
- Regulatory Hurdles
- Accountability

Medical datasets

Medical datasets

- BigBio

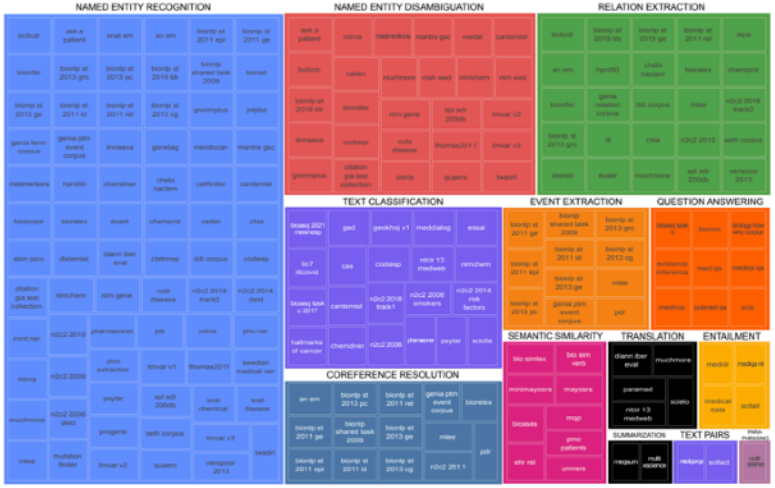


Figure 2: Treemap visualization of BIGBIO's 126 datasets and 13 task categories, denoted by color.

BIGBIO: A Framework for Data-Centric Biomedical Natural Language Processing

Jason Alan Fries[1*] Leon Weber[2,3*] Natasha Seelam[4*] Gabriel Altay[5*]
Debajyoti Datta[6†] Ruisi Su[7†] Samuele Garda[2†] Sunny MS Kang[8†]
Stella Biderman[9,10†] Matthias Samwald[11†] Stephen H. Bach[12†] Wojciech Kusa[13†]
Samuel Cahyawijaya[14†] Fabio Barth[2†] Simon Ott[11†] Mario Sänger[2†] Bo Wang[15]
Alison Callahan[1] Daniel León Periñán[16] Théo Gigant[7] Patrick Haller[2]
Jenny Chim[17] Jose Posada[18] John Giorgi[19] Karthik Rangasai Sivaraman[20]
Marc Pàmies[21] Marianna Nezhurina[22] Robert Martin[2] Moritz Freidank[23]
Nathan Dahlberg[7] Shubhanshu Mishra[24] Shamik Bose[7] Nicholas Broad[25]
Yanis Labrak[26] Shlok S Deshmukh[27] Sid Kiblawi[28] Ayush Singh[7] Minh Chien Vu[29]
Trishala Neeraj[30] Jonas Golde[2] Albert Villanova del Moral[25] Benjamin Beilharz[31]
[1]Stanford University [2]Humboldt-Universität zu Berlin
[3]Max Delbrück Center for Molecular Medicine [4]Sherlock Biosciences [5]Tempus Labs Inc.
[6]University of Virginia [7]BigScience [8]Immuneering [9]EleutherAI [10]Booz Allen Hamilton
[11]Institute of Artificial Intelligence, Medical University of Vienna [12]Brown University
[13]TU Wien [14]The Hong Kong University of Science and Technology
[15–31]See Appendix B *Equal Contribution †Equal Contribution
Corresponding Authors: jason-fries@stanford.edu leonweber@posteo.de
nseelam1@gmail.com gabriel.altay@gmail.com

Medical datasets

- BigBio
- BLURB

Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing

YU GU*, ROBERT TINN*, HAO CHENG*, MICHAEL LUCAS, NAOTO USUYAMA, XIAODONG LIU, TRISTAN NAUMANN, JIANFENG GAO, and HOIFUNG POON, Microsoft Research

BLURB

BLURB is the **B**iomedical **L**anguage **U**nderstanding and **R**easoning **B**enchmark.

DoSSIER MSCA ITN/ETN No. 860721

Medical datasets

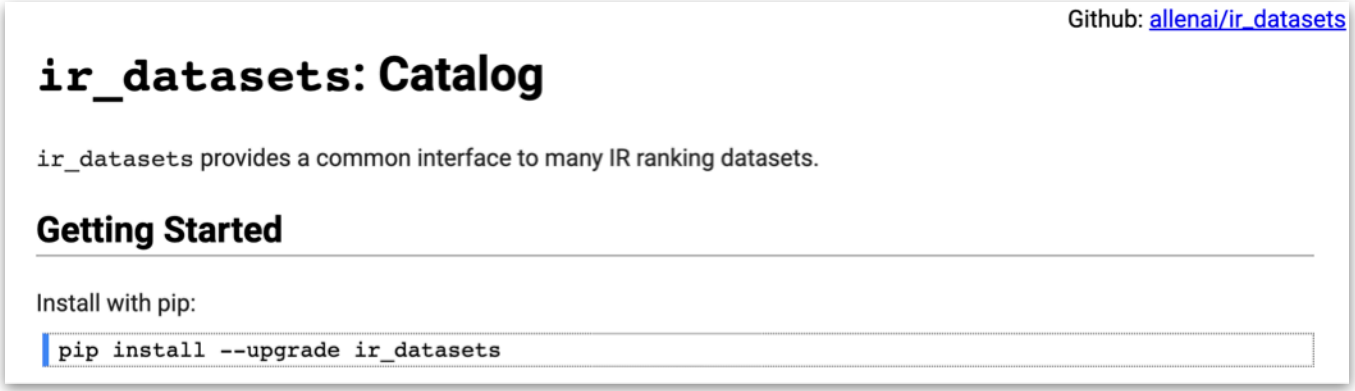
- BigBio
- BLURB
- CBLUE

CBLUE: A Chinese Biomedical Language Understanding Evaluation Benchmark

Ningyu Zhang[1]*, Mosha Chen[2]*, Zhen Bi[1]*, Xiaozhuan Liang[1]*, Lei Li[1]*, Xin Shang[3]
Kangping Yin[2], Chuanqi Tan[2], Jian Xu[2], Fei Huang[2], Luo Si[2], Yuan Ni[4], Guotong Xie[4,5,6]
Zhifang Sui[7,13], Baobao Chang[7,13], Hui Zong[8,14], Zheng Yuan[9], Linfeng Li[10], Jun Yan[10]
Hongying Zan[11,13], Kunli Zhang[11,13], Buzhou Tang[12,13]†, Qingcai Chen[12,13]†
[1]Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies, Zhejiang University
[2]Alibaba Group, [3]School of Mathematical Science, Zhejiang University, [4]Pingan Health Technology
[5]Ping An Health Cloud Company Limited [6]Ping An International Smart City Technology Co., Ltd
[7]Key Laboratory of Computational Linguistics, Ministry of Education, Peking University
[8]School of Life Sciences and Technology, Tongji University [9]Tsinghua University,
[10]Yidu Cloud Technology Inc [11]School of Information Engineering, Zhengzhou University
[12]Harbin Institute of Technology (Shenzhen) [13]Peng Cheng Laboratory, [14]Philips Research China

Medical datasets

- BigBio
- BLURB
- CBLUE
- https://ir-datasets.com



Medical models

- BioBERT
- SciBERT
- BlueBERT
- ClinicalBERT

- ClinicalLongformer
- ClinicalBigBird
- …

Agenda

1. Neural retrieval recap
2. Challenges in medical neural retrieval
3. Neural retrieval for clinical trials
4. Introduction to the TCRR model
5. Performance tracking
6. Hands-on #1: model training
7. Hands-on #2: inference

Neural retrieval for clinical trials

1. Neural query expansion with T5
2. Dense retrieval
3. Cross encoder

DoSSIER Domain Specific Systems for Information Extraction and Retrieval MSCA ITN/ETN No. 860721

SHORT-PAPER

Neural Query Synthesis and Domain-Specific Ranking Templates for Multi-Stage Clinical Trial Matching

Authors: Ronak Pradeep, Yilin Li, Yuetong Wang, Jimmy Lin Authors Info & Claims

SIGIR '22: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval • July 2022 • Pages 2325–2330 • https://doi.org/10.1145/3477495.3531853

Published: 07 July 2022 Publication History

Check for updates

ABSTRACT

In this work, we propose an effective multi-stage neural ranking system for the clinical trial matching problem. First, we introduce NQS, a neural query synthesis method that leverages a zero-shot document expansion model to generate multiple sentence-long queries from lengthy patient descriptions. These queries are independently issued to a search engine and the results are fused. We find that on the TREC 2021 Clinical Trials Track, this method outperforms strong traditional baselines like BM25 and BM25 + RM3 by about 12 points in nDCG@10, a relative improvement of 34%. This simple method is so effective that even a state-of-the-art neural relevance ranking method trained on the medical subset of MS MARCO passage, when reranking the results of NQS, fails to improve on the ranked list. Second, we introduce a two-stage neural reranking pipeline trained on clinical trial matching data using tailored ranking templates. In this setting, we can train a pointwise reranker using just 1.1k positive examples and obtain effectiveness improvements over NQS by 24 points. This end-to-end multi-stage system demonstrates a 20% relative effectiveness gain compared to the second-best submission at TREC 2021, making it an important step towards better automated clinical trial matching.

DoSSIER MSCA ITN/ETN No. 860721

Journal of Biomedical Informatics
Volume 144, August 2023, 104444

Effective matching of patients to clinical trials using entity extraction and neural re-ranking

Wojciech Kusa [a 1], Óscar E. Mendoza [c 1], Petr Knoth [d], Gabriella Pasi [c], Allan Hanbury [a b]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.jbi.2023.104444

Get rights and content

Highlights

- We conduct several experiments for the patient-to-trial matching retrieval problem.
- Inclusion criteria section has the biggest influence on the score in lexical models.
- Query and document enrichment techniques improve retrieval of relevant trials.
- Age and gender-based filtering helped remove 26% ineligible trials.
- Novel training strategy for re-ranking further increases retrieval effectiveness.

DoSSIER
MSCA ITN/ETN No. 860721

Matching a Patient from An Admission Note to Clinical Trials: Experiments with Query Generation and Neural-Ranking

Vincent Nguyen Maciej Rybinski Sarvnaz Karimi
CSIRO Data61
Sydney, Australia
firstname.lastname@csiro.au

ABSTRACT

Many clinical trials fail to attract enough eligible participants. The TREC 2022 Clinical Trials track set a task where patient data, in the form of clinical notes, can be used to match eligible patients to a relevant clinical trial. We explore a number of dense retrieval methods using Bidirectional Encoder Representations from Transformers (BERT). Our best method used BERT reranking using models based on monoBERT architecture. Our self-supervised monoBERT run achieved effectiveness competitive to that of a fully-tuned monoBERT run.

CCS CONCEPTS

• **Information systems** → **Retrieval models and ranking**; *Language models*; *Decision support systems*; • **Applied computing** → *Health informatics*.

KEYWORDS

Clinical trials search; Medical information retrieval; Learning-to-rank; Evidence-based medicine

ACM Reference Format:
Vincent Nguyen Maciej Rybinski Sarvnaz Karimi. 2021. Matching a Patient from An Admission Note to Clinical Trials: Experiments with Query Generation and Neural-Ranking. In *TREC'22: TREC, November, 2022*. ACM, New York, NY, USA, 4 pages.

tasks [4–6] in its 2017, 2018, and 2019 editions. The task of this year's track is to link a synthetic patient's electronic health record (EHR), in free text, to relevant clinical trials. TREC CT's goal is to study the use of automatic retrieval systems to expose patients to relevant clinical trials to increase participation.

In our submission to this year in the TREC CT track, we build upon our last year's submission [7]. Our experiments this year focus on neural ranking using resource-effective self-supervision and supervision signals from last year's judgement pool. Our experiments with a neural reranking pipeline centered around resource-effective learning, used a reranker trained on labeled data (from last year's edition of the track) compared with a self-supervised model trained using the target document corpus. We also experiment with efficient end-to-end neural ranking (where document representations can be pre-computed) with bi-encoders and with neural query expansion. Finally, we also probe the effect of a simple heuristic for matching the patient note with the demographic profile specified in the clinical trials, which we apply to one of the bi-encoder runs.

2 DATASET

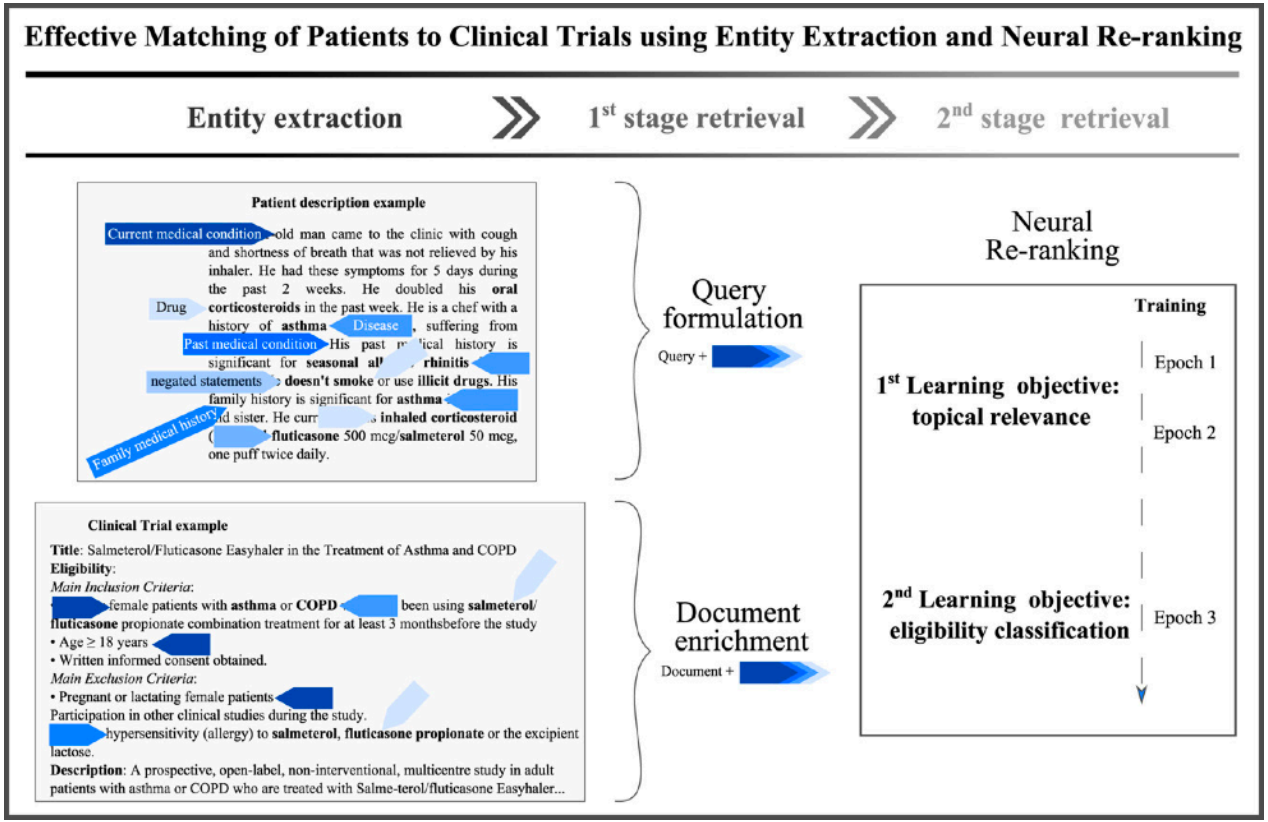
The TREC 2022 CT dataset consists of 50 topics with 35,394 relevance judgments. The corpus for the task is a 2020 snap-

DoSSIER MSCA ITN/ETN No. 860721

Agenda

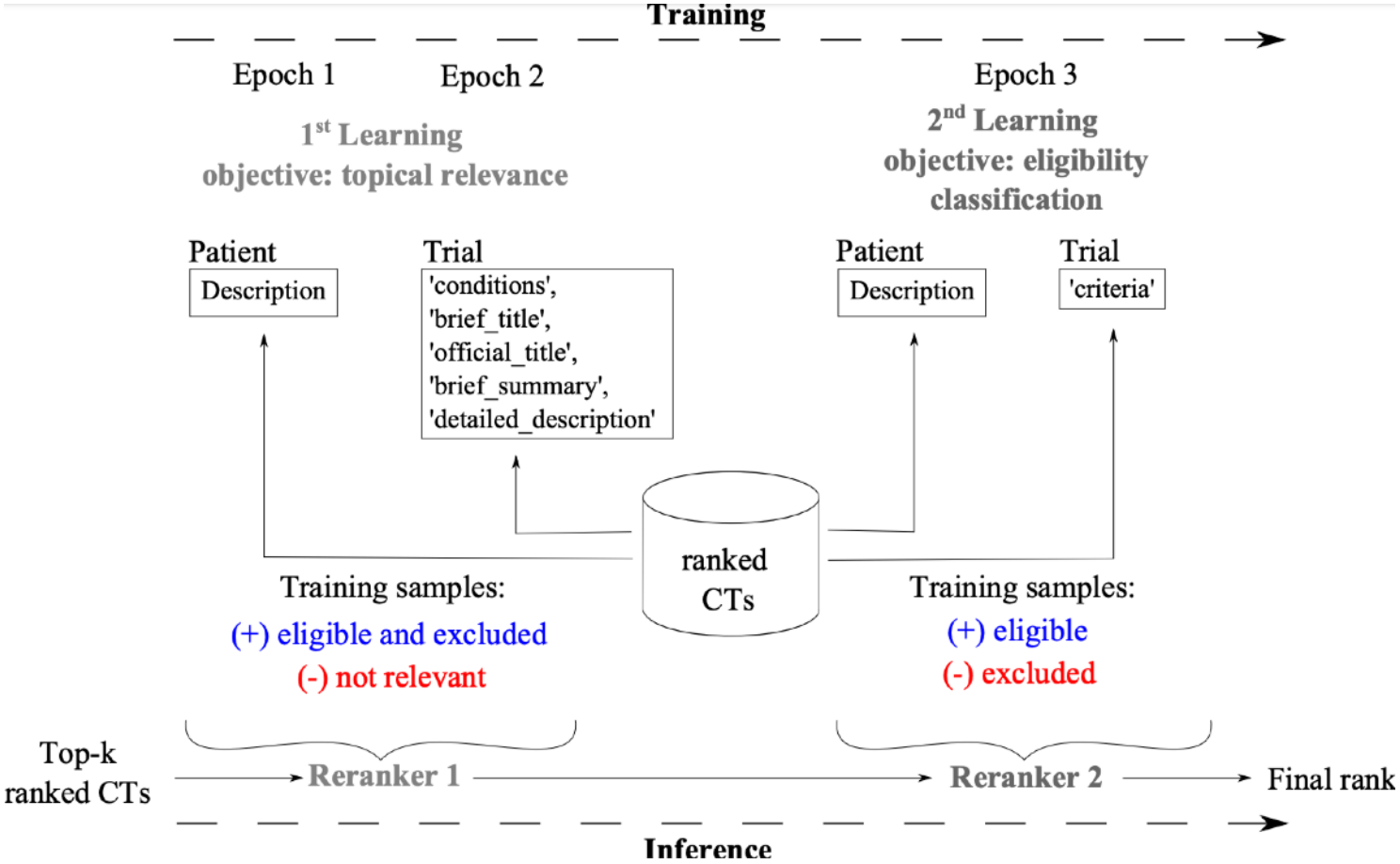
1. Neural retrieval recap
2. Challenges in medical neural retrieval
3. Neural retrieval for clinical trials
4. Introduction to the TCRR model
5. Performance tracking
6. Hands-on #1: model training
7. Hands-on #2: inference

Topical and Criteria Re-Ranking — TCRR



DoSSIER
MSCA ITN/ETN No. 860721

Training Pipeline



DoSSIER MSCA ITN/ETN No. 860721

Agenda

1. Neural retrieval recap
2. Challenges in medical neural retrieval
3. Neural retrieval for clinical trials
4. Introduction to the TCRR model
5. Performance tracking
6. Hands-on #1: model training
7. Hands-on #2: inference

What can be visualised?

1. Model architecture
2. Model training
3. Model inference

What can be visualised?

1. Model architecture
2. Model training
3. Model inference

Wandb example

```python
# train.py
import wandb
import random # for demo script

wandb.login()

epochs=10
lr=0.01

run = wandb.init(
    # Set the project where this run will be logged
    project="my-awesome-project",
    # Track hyperparameters and run metadata
    config={
        "learning_rate": lr,
        "epochs": epochs,
    })

offset = random.random() / 5
print(f"lr: {lr}")

# simulating a training run
for epoch in range(2, epochs):
    acc = 1 - 2 ** -epoch - random.random() / epoch -
offset
    loss = 2 ** -epoch + random.random() / epoch + offset
    print(f"epoch={epoch}, accuracy={acc}, loss={loss}")
    wandb.log({"accuracy": acc, "loss": loss})

# run.log_code()
```

DoSSIER MSCA ITN/ETN No. 860721

Wandb – PyTorch example

```python
import wandb
wandb.init(config=args)

# set up your model
model = AutoModelForSequenceClassification.from_pretrained(
                model_name, num_labels=2
            )

# Magic
wandb.watch(model, log_freq=100)

model.train()
for batch_idx, (data, target) in enumerate(train_loader):
    output = model(data)
    loss = F.nll_loss(output, target)
    loss.backward()
    optimizer.step()
    if batch_idx % args.log_interval == 0:
        wandb.log({"loss": loss})
```

Wandb — logging metrics

```python
predictions = trainer.predict(
    tokenized_datasets[test_data],
)

wandb.log(
    {
        f"{test_data}_precision": predictions.metrics["test_precision"],
        f"{test_data}_recall": predictions.metrics["test_recall"],
        f"{test_data}_f1": predictions.metrics["test_f1"],
        f"{test_data}_loss": predictions.metrics["test_loss"],
    }
)
```

Agenda

1. Neural retrieval recap
2. Challenges in medical neural retrieval
3. Neural retrieval for clinical trials
4. Introduction to the TCRR model
5. Performance tracking
6. Hands-on #1: model training
7. Hands-on #2: inference

Hands-on Session 1: Training with TCRR

- Google Colab

Agenda

1. Neural retrieval recap
2. Challenges in medical neural retrieval
3. Neural retrieval for clinical trials
4. Introduction to the TCRR model
5. Performance tracking
6. Hands-on #1: model training
7. Hands-on #2: inference

Hands-on Session 2: Inference and Application

- Google Colab

Questions?