



ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

CS-433 MACHINE LEARNING

PROJECT 2 – FALL 2025

---

# **Date Prediction for Museum Collections Using BERT Embeddings**

---

IN COLLABORATION WITH

**Laboratory for Experimental Museology (eM+)**

<b>Louis Larcher</b>	SCIPER: 361002
<b>Cassio Manueguerra</b>	SCIPER: 346232
<b>Arthur Taieb</b>	SCIPER: 361195

17th December 2025

## ABSTRACT

This report presents our work on predicting creation dates for artworks in the Wellcome Collection, conducted in collaboration with the Laboratory for Experimental Museology (eM+) at EPFL. Given the incomplete metadata in museum digital archives, automated date prediction offers significant value for collection curation and research. We explore two approaches: (1) using frozen BERT embeddings combined with a linear classifier, and (2) fine-tuning BERT end-to-end with the classification head. Our methods leverage textual descriptions, titles, and other metadata fields to predict temporal categories. We evaluate both approaches and discuss their trade-offs in terms of accuracy and computational requirements.

# 1 INTRODUCTION

## 1.1 CONTEXT AND MOTIVATION

Museum collections worldwide are undergoing large-scale digitization efforts, making cultural heritage more accessible than ever. The Wellcome Collection [1] is one such open-access repository containing hundreds of thousands of artworks, manuscripts, and artifacts related to the history of medicine and human experience. However, like many historical collections, metadata completeness varies significantly-particularly regarding precise dating of objects. Accurate dating of museum artifacts is crucial for art historical research, exhibition curation, and understanding the evolution of visual and material culture. Manual annotation by experts is time-consuming and not scalable to large collections. This motivates the development of automated methods that can leverage existing textual metadata to predict missing dates.

## 1.2 COLLABORATION WITH EM+

This project is conducted in collaboration with the Laboratory for Experimental Museology (eM+) at EPFL. Initial discussions with the lab explored various potential contributions, including:

- Visualization of collection embeddings for exploratory analysis
- Recommendation systems for similar artworks

However, based on feedback from eM+ researchers, we pivoted to focus on **date prediction**, as this addresses a more immediate and practical need for their work.

## 1.3 PROBLEM STATEMENT

Given textual metadata fields (titles, descriptions, physical description, contributors, etc.) associated with artworks in the Wellcome Collection, our objective is to predict the creation date or date range of each item. We formulate this as a classification task, where the target classes represent temporal periods or date ranges.

# 2 METHODS

**Global approach.** We explore two distinct approaches for our task, each then also divided in two subcases: regression and classification.

**Frozen BERT with a trained head:** In this approach, a pre-trained BERT model acts as a fixed feature extractor. The BERT layers are frozen, and only a custom classification or regression head (e.g., a linear layer) is trained on top of the extracted embeddings.

**Unfrozen (Fine-tuned) BERT with a trained head:** This approach involves fine-tuning the entire BERT model along with its custom head. Backpropagation flows through all layers of the model, allowing BERT’s internal representations to adapt to our specific dataset and task.

## 2.1 DATA AND PREPROCESSING

The Wellcome Collection provides open access to its catalog data. Our dataset consists of artwork records containing various textual fields like title, worktype, description, languages, contributors, physical description, etc.

Only title and worktype are always available, while the rest may be missing in some records. For example, the target date field is not available for only TODO% of the records [2].

**Text Merging.** As our model is based on BERT, we need every text field into one. Therefore we concatenate all available text fields into a single input string, that follows recommended patterns to help BERT better understand the context and the fields name meaning.

**Data Cleaning.** We performed several data cleaning steps to prepare the dataset for model training. First, we removed non-meaningful or near fully empty columns such as various identification numbers, lettering or edition. Next, we removed all records with missing dates from our dataset. We identified and removed outliers in the date field, it gives us a dataset of TODO records with date fields going from TODO to TODO.

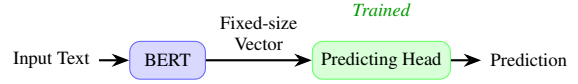
**Date Preprocessing.** Before everything to avoid any dataleak we can now split our dataset into train and test sets. Then, depending on the specific task, we processed the date information differently: for classification tasks, we one-hot encoded relevant time periods, while for regression tasks, we normalized the dates to a continuous range between 0 and 1.

## 2.2 APPROACH 1: FROZEN BERT

Our first approach uses pre-trained BERT [3] as a fixed feature extractor. The model weights remain unchanged during training, allowing us to benefit from BERT’s pre-trained language understanding without the computational cost of fine-tuning.

**Architecture.** The input text is encoded by BERT into a fixed-size vector, then passed through a predicting head to produce the final prediction. We try different head architectures to compare their performance, both are

multi-layer perceptrons (MLP) once use for regression and once for classification.



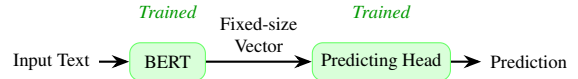
**FIGURE 1**  
Frozen BERT architecture

**Training Strategy.** Since only the head is trainable, we can use a higher learning rate (e.g. TODO). Training is fast and memory-efficient, making this approach suitable for rapid experimentation.

## 2.3 APPROACH 2: BERT

Our second approach allows the BERT model to adapt its representations by training end-to-end.

**Architecture.** The architecture is the same as in Approach 1, but now the entire model is trainable.



**FIGURE 2**  
Fine-tuned BERT architecture

**Training Strategy.**

- Lower LR for BERT (e.g.,  $2e-5$ )
- Higher LR for head
- Warmup schedule
- Early stopping

**Computational Considerations.** Fine-tuning is significantly more expensive, requiring GPU acceleration for this we used Kaggle.

## 2.4 EVALUATION METRICS

We evaluate both approaches using:

- **Accuracy:** Overall accuracy
- **Macro F1:** Class-balanced score
- **Confusion Matrix:** Per-class analysis

# 3 RESULTS

## 3.1 EXPERIMENTAL SETUP

All experiments were conducted using PyTorch with the HuggingFace Transformers library, using the `bert-base-uncased` model as our backbone.

## 3.2 MODEL COMPARISON

### FROZEN BERT + LINEAR

### FINE-TUNED BERT

### 3.3 SUMMARY TABLE

**TABLE 1**  
Comparison of Date Prediction Approaches

Method	Accuracy	Macro F1	Time
Frozen BERT	[-]	[-]	[-]
Fine-tuned BERT	[-]	[-]	[-]

### 3.4 ANALYSIS

#### KEY FINDINGS

- Finding 1: [TO BE FILLED]
- Finding 2: [TO BE FILLED]
- Finding 3: [TO BE FILLED]

#### TRADE-OFFS

- Frozen: faster training, lower memory
- Fine-tuned: better accuracy, higher cost

## 4 CONCLUSION

### 4.1 SUMMARY

In this project, we addressed the challenge of automatically predicting creation dates for artworks in the Wellcome Collection, in collaboration with the Laboratory for Experimental Museology (eM+). We developed and compared two approaches based on BERT language models:

1. **Frozen BERT:** Using pre-trained embeddings with a trainable linear classifier, offering computational efficiency while achieving [TO BE FILLED] accuracy.
2. **Fine-tuned BERT:** End-to-end training that adapts the model to the date prediction task, achieving [TO BE FILLED] accuracy at higher computational cost.

Our results demonstrate that textual metadata in museum collections contains sufficient signal for temporal prediction, and that modern language models can effectively capture this information.

### 4.2 LIMITATIONS

Several limitations should be acknowledged:

- The date categories used may not align perfectly with art historical periods
- Ambiguous or approximate dates in the ground truth affect model evaluation
- Our experiments focused on English-language metadata only
- The single-collection focus may limit generalizability

### 4.3 FUTURE WORK

Promising directions for future research include:

- **Multimodal approaches:** Incorporating image data alongside text for improved predictions
- **Regression formulation:** Predicting continuous year values rather than categorical periods
- **Cross-collection transfer:** Testing generalization to other museum collections
- **Uncertainty quantification:** Providing confidence estimates for predictions
- **Integration with eM+ tools:** Deploying the model in practical museum curation workflows

### 4.4 ACKNOWLEDGMENTS

We thank the Laboratory for Experimental Museology (eM+) at EPFL for their guidance and domain expertise, and the Wellcome Collection for making their data openly available.

## REFERENCES

- [1] Wellcome Trust. *Wellcome Collection*. <https://wellcome.org/>. Accessed: 2024. 2024.
- [2] Louis Larcher, Cassio Manueguerra and Arthur Taieb. *Missing Values Analysis for Wellcome Collection Dataset*. *Images/NaN\_percentage.png*. Statistical analysis of data completeness. 2025.
- [3] Jacob Devlin et al. ‘BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding’. In: *Proceedings of NAACL-HLT* (2019), pp. 4171–4186.