Transfer Learning, Pretraining, Finetuning

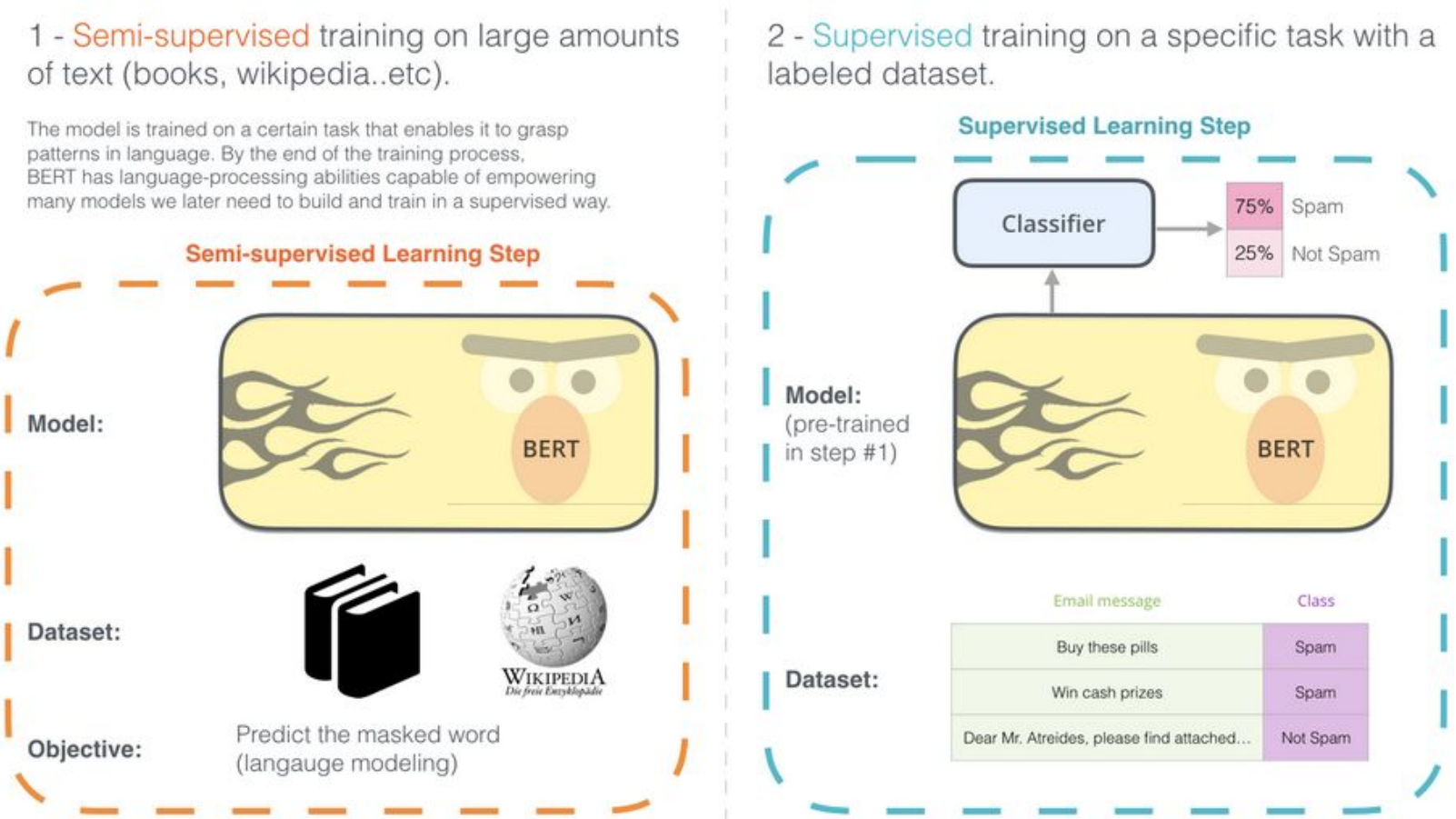
Transfer Learning

- Transfer knowledge from one task towards another task
- You do it too: The first programming language you learn is probably the hardest

- Reason: related tasks require similar sets of skills

Transfer Learning in NLP

- All text-based NLP tasks require some representations of text
- But building representations for each task would be a waste of time
- Solution: Start with general text representations, adapt for specialized tasks
- already used this method with static word embeddings, now take it further



The two steps of how BERT is developed. You can download the model pre-trained in step 1 (trained on un-annotated data), and only worry about fine-tuning it for step 2. [Source for book icon].

IMG: https://jalammar.github.io/illustrated-bert/

Crosslingual transfer

- Most languages have less resources than others
- But related languages share vocabulary, grammatical structure etc.
- Ex.: Knowledge of French, Spanish, Portuguese will help you to model Italian
- Mutlilingual LMs: Multilingual BERT, XLM-RoBERTa (Conneau et al. 2020)
- You probably experienced crosslingual transfer when learning languages!

Practical Session: Fine-tuning a LM

- Select one of the fine-tuning notebooks from this repo: https://github.com/huggingface/transformers/tree/master/notebooks
- E.g. “How to fine-tune a model on token classification” for POS/NER.
- Run it in Google Colab. You can do that without a huggingface account & token. Skip the notebook_login() cell. When building the TrainingArguments object, set `push_to_hub=False`