

## Mini-Challenge 2: Paper-Studium und Umsetzung

### Ziel

Vertiefung in ein eher aktuelles Paper aus der Forschung und Umsetzung eines darin beschriebenen oder verwandten Tasks - gemäss Vereinbarung mit dem Fachcoach.

*Beispiel:* Implementiere, trainiere und validiere ein Deep Learning Modell für Image Captioning wie beschrieben z.B. im Paper Show and Tell.

### Arbeitsschritte

#### Schritt 1: Paper

Wahl des Papers.

**Absprache/Beschluss** mit Coach und Beschluss, welches Paper genommen wird.

#### Schritt 2: Aufbau Modellierung und Daten

Überlege Dir, welche Modell-Architektur Sinn machen könnte. Mindestens zwei Modell-Varianten sollen aufgebaut werden, die miteinander verglichen werden sollen.

Für das Captioning-Beispiel kann z.B. das Paper von [Vinyals](#) oder, für einen etwas abgewandelten Task, auch das Paper von [Karpathy](#) herangezogen werden. Die zwei Modell-Varianten könnten hier darin bestehen, dass neben dem im gewählten Paper beschriebenen Modell auch ein vortrainiertes Modell (z.B. ein Foundation-Modell) herangezogen wird. Auch bei den Daten bestehen Wahlmöglichkeiten. Allerdings sollte darauf geachtet werden, dass ein ML-ready Datenset verwendet wird, so dass sich der Aufwand fürs Daten-Preprocessing in Grenzen hält. Für den Captioning Task könnte beispielsweise [Flickr 8k](#) oder auch [RSICD](#) verwendet werden.

**Absprache/Beschluss** mit Coach und Beschluss, was evaluiert werden soll.

#### Schritt 3: Training und Evaluation

Trainiere und evaluiere das Modell. Beschreibe genau was Du tust und warum Du es tust.

#### Schritt 4: Präsentation / Bericht

**Präsentation** (~10m): Kurze Präsentation mit Diskussion der wichtigsten Ergebnisse.

**Q&A** (~10min): Klärung von Verständnisfragen zum Paper und der Umsetzung.

**Bericht** in Form **eines** gut dokumentierten, übersichtlichen Jupyter Notebooks.

Dieses soll schliesslich auch abgegeben werden und dem Fachexperten erlauben, die Schritte nachzuvollziehen (allenfalls auch das Training erneut laufen zu lassen).

### Zeitlicher Rahmen:

Wird beim Schritt 1 **verbindlich** festgelegt.

## Beurteilung

Beurteilt wird auf Basis des abgegebenen Notebooks:

- Vollständige und korrekte Umsetzung der vereinbarten Aufgabestellung.
- Klare, gut-strukturierte Umsetzung.
- Schlüssige Beschreibung und Interpretation der Ergebnisse. Gut gewählte und gut kommentierten Plots und Tabellen.
- Vernünftiger Umgang mit (Computing-)Ressourcen.
- Verständliche Präsentation der Ergebnisse.

## Referenzen, Key Words

- Word Embedding (z.B. word2vec, glove), um Wörter in numerische Vektoren in einem geeignet dimensionierten Raum zu mappen.  
Siehe z.B. Andrew Ng, Coursera: <https://www.coursera.org/lecture/nlp-sequence-models/learning-word-embeddings-APM5s>
- Bild Embedding mittels vortrainierten (evt. retrained) Netzwerken wie beispielsweise ResNet, GoogLeNet, EfficientNet oder ähnlich. Transfer-Learning.
- Seq2Seq Models bekannt für Sprach-Übersetzung.