

## Mini-Challenge 2: Image Captioning

### Ziel

Implementiere, trainiere und validiere ein Deep Learning Modell für Image Captioning.

### Arbeitsschritte

#### Schritt 1: Daten

[https://ivc.ischool.utexas.edu/VizWiz\\_visualization](https://ivc.ischool.utexas.edu/VizWiz_visualization)

#### Schritt 2: Aufbau Modellierung

Überlege Dir, welche Modell-Architektur Sinn machen könnte. Das Paper von Vinyals (<https://arxiv.org/pdf/1411.4555.pdf>) war wichtig in diesem Zusammenhang. du kannst Dich auf Ihre Architektur beziehen oder auch eine eigene, evt. einfachere verwenden.

Zwei Modell-Varianten sollen aufgebaut werden:

- Ohne Transfer-Learning für den Bild-Teil
- Mit Transfer-Learning für den Bild-Teil

#### Schritt 3: Training und Evaluation

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

#### Schritt 4: Bericht

Test mit eigenen Bildern.

### 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 oder ähnlich. Transfer-Learning.
- Seq2Seq Models bekannt für Sprach-Übersetzung.

### Abgabe

Code und Bericht.

**Deadline:** Gemäss Vereinbarung.

## Beurteilung

Beurteilt wird auf Basis des abgegebenen Notebooks:

- Vollständige und korrekte Umsetzung der oben beschriebenen Aufgabestellung (Implementierung im DL-Framework Deiner Wahl, Untersuchung der verschiedenen Varianten).
- Gut gewählte und gut kommentierten Plots und Tabellen.
- Schlüssige Beschreibung und Interpretation der Ergebnisse.