

# TiU-Talpa Internships



# What is this program about: TiU point of view

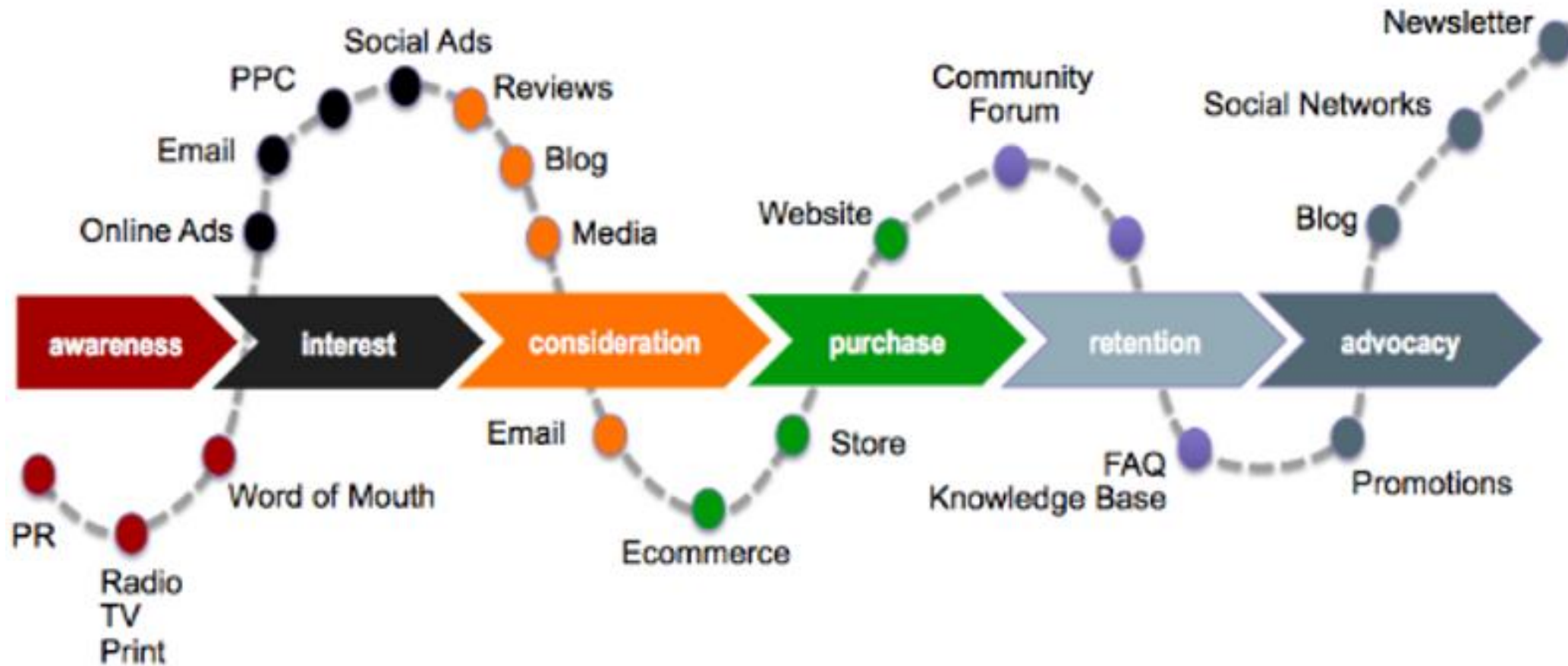


---

Understanding Society

**Digital Dutch Society**

# Digital Journey



# Collaborations between Talpa and TiU

## Data Inspired Creativity

*Leveraging the digital Journey for producing better media content*

As part of NWO-Creative industry collaborations

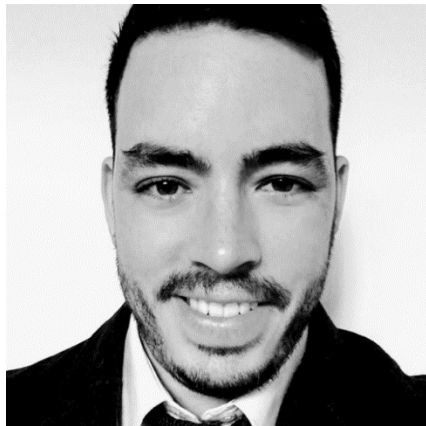
More info at:

<https://www.tilburguniversity.edu/about/schools/economics-and-management/organization/nwo-grant-research-project-data-inspired-creativity-collaboration-talpa-media>

# But let's talk about Internships

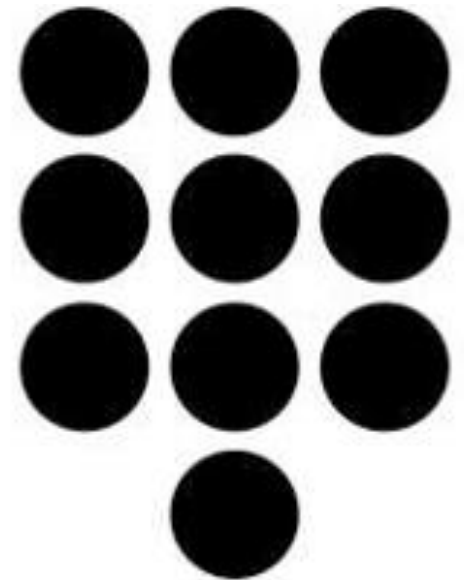
Dr. Anca Dumitrache

<https://www.linkedin.com/in/dumitracheanca/>



Ricardo Fabian Guevara

<https://www.linkedin.com/in/ricardo-fabian-guevara/>



# Do you know Talpa?

Let's use the words of Wouter Postma, Vice president at Talpa

[https://www.linkedin.com/posts/wouterpostma\\_innovation-technology-digital-activity-6532146983698788352-ZkUw/](https://www.linkedin.com/posts/wouterpostma_innovation-technology-digital-activity-6532146983698788352-ZkUw/)

More info at: <https://talpanetwork.com/>

Current open positions: <https://jobs.talpanetwork.com/>

# Talpa Internships Opportunities

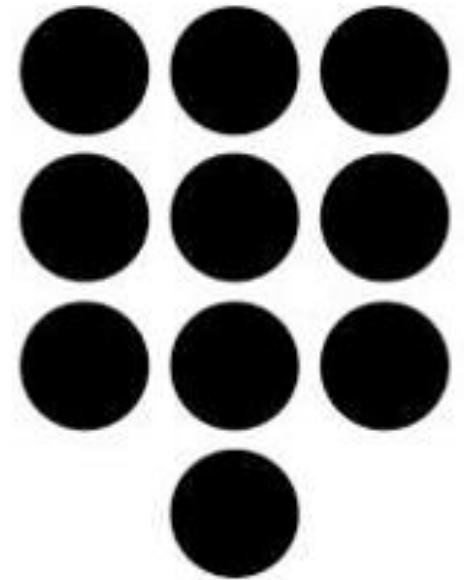
Dr. Anca Dumitrache

<https://www.linkedin.com/in/dumitracheanca/>



Ricardo Fabian Guevara

<https://www.linkedin.com/in/ricardo-fabian-guevara/>



# Who is your ideal candidate?

For our student, check the Webinar at the following link:

<https://videocollege.uvt.nl/Mediasite/Play/767355b24a5e4307811f59f802f1c04c1d>



# Swimlane for Trending on Social Media

- Both KIJK (video streaming platform) and JUKE (audio streaming platform) present their content separated into swimlanes on the front page. For instance, one swimlane contains a list of TV shows that were popular in the previous days. We would like to create a new swimlane that contains items that were trending on social media. Taking Twitter as a data source, the project will go through the following steps: (1) extract tweets about popular TV shows and/or radio shows, (2) perform entity linking to match them to the shows in our database, (3) aggregate the results to get the most popular shows in one swimlane.
- *Project:* KIJK and/or JUKE

# Automatic Teaser Tweet Creation

- Given a (textual) description of a TV show episode or radio program, we would like to generate teaser tweets about the show that are meant to generate anticipation on social media. The underlying task would be a summarization problem, where the program description is mapped to a short tweet about it. The tweet should contain relevant information about the program, but not reveal any spoilers. Either an extractive (entity + relation extraction) or abstractive method could be applied.
- *Source:* <https://www.aclweb.org/anthology/N19-1398/>
- *Project:* KIJK and/or JUKE

# Exploring Pair-wise Learning to Rank

- Our current recommender system uses ALS, a point-wise learning to rank approach, where the learning objective is based on modeling the score of a given item (i.e. similarly to how regression works). Alternative methods of doing learning-to-rank are pair-wise (learning objective is to model the ranking of a pair of items relative to each other) and list-wise learning to rank (learning objective is calculated over the entire list of items). The project goal is to investigate different learning objectives and find out: (1) how they perform relatively to the point-wise method, (2) if there are subsets of data where this method applies better/worse.
- *Source:* <https://medium.com/@nikhilbd/pointwise-vs-pairwise-vs-listwise-learning-to-rank-80a8fe8faddf>
- *Project:* KIJK and/or JUKE

# Automatic Playlist Generation

- JUKE music player features a lot of "non-stop music" playlists that are manually created by an editor, usually selecting music from a given genre (e.g. hard rock non-stop radio). We would like to see whether these playlists can be generated by AI and what the quality is. This can be approached as a song clustering problem, where the feature space could contain the genre, artist, as well as other audio features.
- Project: JUKE



# Google AdWords for video

- Goal: Advertisers can "buy" Google search keywords to show their ad with. We would like to see whether this is possible with keywords/topics within videos too, so contextual advertising becomes feasible. The steps involved in this project are: (1) generate textual metadata of video (either based on transcript, or other features of the video; there might be some existing video metadata as well), (2) match video textual metadata with ad keywords.

# Exploring Seasonal Trends as RecSys Features

- Goal: produce embeddings with various sizes similar to Glove but for radio stations, that reflect radio station similarity with euclidean distance.
- Value: many of our problems could benefit, like radio to radio similarity. Our RecSys can also use the radio embedding as a feature to determine if a radio is a good match for a user. This could also greatly alleviate the cold start problem when we introduce new radio stations.
- Approach: this is an unexplored problem but there has been work on audio embeddings. In the simplest form, a radio station embedding could be just a bag of features like genre frequencies, languages and origin of artists, but also properties from the target audience like age range. Going beyond that, it could use an average of the embeddings of representative songs on it in a given past period. Another valuable asset from Talpa is the usage behaviour, so radio station features could also be calculated based on the type of person who listens to it, the frequency and other seasonal behaviour.
  - There is a time aspect to radio stations so we cannot expect the embeddings to remain the same forever. A refresh period is to be expected (even word embeddings need to account for word semantic drifting, only their refresh window is larger).

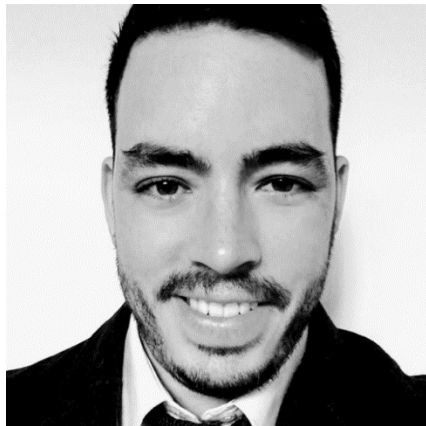
# Podcast Embeddings for JUKE

- Similar to the radio embeddings project, we are interested in creating embeddings for the podcasts available on JUKE. Podcast recommendations suffer from cold start problem even more than radio, since there is usually a high volume of items that are published continuously. The features that can be used for podcast embeddings are also slightly different than typical audio embedding features based on e.g. musical genre.

# Q&A

Dr. Anca Dumitrache

<https://www.linkedin.com/in/dumitracheanca/>



Ricardo Fabian Guevara

<https://www.linkedin.com/in/ricardo-fabian-guevara/>

