A close-up of a logo

Description automatically generated

**Automatic alert generation with NER and SA**

**Ulises Díez Santaolalla**

**Teresa Franco Corzo**

**Ignacio Felices Vera**

**Grupo A**

**Deep Learning - Natural Language Processing**

**3º Grado en Ingeniería Matemática e Inteligencia Artificial**

Contenido

[Project Plan and Milestones 3](#_Toc193828590)

[Phase 1: Research and Data Collection 4](#_Toc193828591)

## Project Plan and Milestones

**Phase 1: Research and Data Collection (Weeks 1-4)**

* Conduct an in-depth literature review.
* Collect news articles and social media posts.
* Gather image datasets with captions (e.g., COCO, Flickr30k).
* Preprocess textual and visual data.

**Phase 2: Model Selection and Preprocessing (Weeks 5-8)**

* Implement and fine-tune a custom LSTM-CRF NER model.
* Train/test a sentiment analysis model using CNN or LSTM.
* Utilize a pre-trained image captioning model (e.g., BLIP) and fine-tune it.

**Phase 3: Multi-Modal Integration (Weeks 9-12)**

* Develop a mechanism to integrate textual and visual data.
* Experiment with different architectures to combine NER, SA, and image captioning outputs.
* Implement a Seq2Seq model for alert generation.

**Phase 4: Evaluation and Optimization (Weeks 13-16)**

* Evaluate each component using standard benchmarks (F1-score, BLEU, ROUGE, etc.).
* Optimize models through hyperparameter tuning.
* Conduct ablation studies to assess the impact of each component.

**Phase 5: Documentation (Weeks 17-20)**

* Conduct final testing and prepare for deployment.
* Write comprehensive documentation and final report.

## Phase 1: Research and Data Collection

**1. Named Entity Recognition (NER)**

Named Entity Recognition (NER) is essential for extracting structured information from unstructured text. Conventional models use Conditional Random Fields (CRFs) and Hidden Markov Models (HMMs), but deep learning approaches such as Long Short-Term Memory (LSTM) networks with Conditional Random Fields (LSTM-CRF) and Transformer-based architectures (e.g., BERT-based models) have shown superior performance.

**Bibliography**

* Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). "Neural Architectures for Named Entity Recognition."

<https://arxiv.org/abs/1603.01360>

* Yadav, V., & Bethard, S. (2018). "A Survey on Recent Advances in Named Entity Recognition from Deep Learning Models.".

<https://arxiv.org/abs/1910.11470>

**2. Sentiment Analysis (SA)**

Sentiment Analysis (SA) classifies text into predefined sentiment categories (e.g., positive, negative, neutral). Classical methods rely on lexicon-based and machine learning approaches (SVM, Naïve Bayes), but modern methods utilize deep learning architectures such as CNNs, LSTMs, and Transformers.

**Bibliography**

* Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank."

<https://aclanthology.org/D13-1170/>

* Zhang, L., Wang, S., & Liu, B. (2018). "Deep Learning for Sentiment Analysis: A Survey."

<https://arxiv.org/abs/1801.07883>

**3. Image Captioning for Multi-Modal Processing**

Image captioning is crucial for integrating visual and textual data. State-of-the-art models employ a CNN-RNN architecture, where CNNs extract image features and RNNs generate descriptive captions. Transformer-based approaches (e.g., BLIP, OFA) have further improved performance.

**Bibliography**

* Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). "Show and Tell: A Neural Image Caption Generator."

<https://arxiv.org/abs/1411.4555>

* Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., Zemel, R., & Bengio, Y. (2015). "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention."

<https://arxiv.org/abs/1502.03044>

**4. Alert Generation (AG)**

Combining NER and SA outputs into meaningful alerts requires sequence-to-sequence (Seq2Seq) architectures, often incorporating attention mechanisms or Transformer-based models. Fine-tuning pre-trained models on domain-specific data can improve accuracy.

**Bibliography**

* Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). "Attention is All You Need."

<https://arxiv.org/abs/1706.03762>

* Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2020). "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension."

<https://arxiv.org/abs/1910.13461>