Towards an inclusive Metaverse

Luís Gonzaga Mendes Magalhães

CCG/ZDGV Institute

Computer Vision, Interaction and Graphics Department

lmagalhaes@dsi.uminho.pt

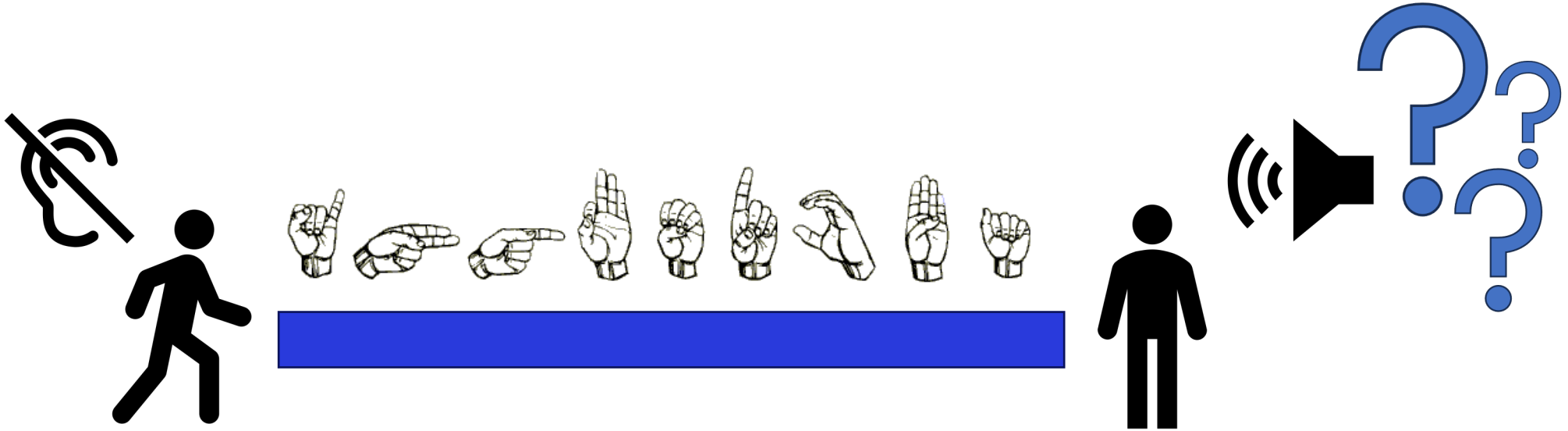
Web3D 2023
SAN SEBASTIAN - SPAIN

Overview

- Problem contextualization – inclusive communication for deaf and hearing people
- IVLing project
- Portuguese sign recognition solution proposal
- System overview
- System in practice
- Results
- Conclusions and future work

Problem contextualization

- Deaf community performs communication through Sign Language
- It combines gestures involving hand configurations and movements, facial expressions and even body posture
- The communication between a deaf and a hearing person rarely occurs without mediation, usually, an interpreter



IVLinG - project

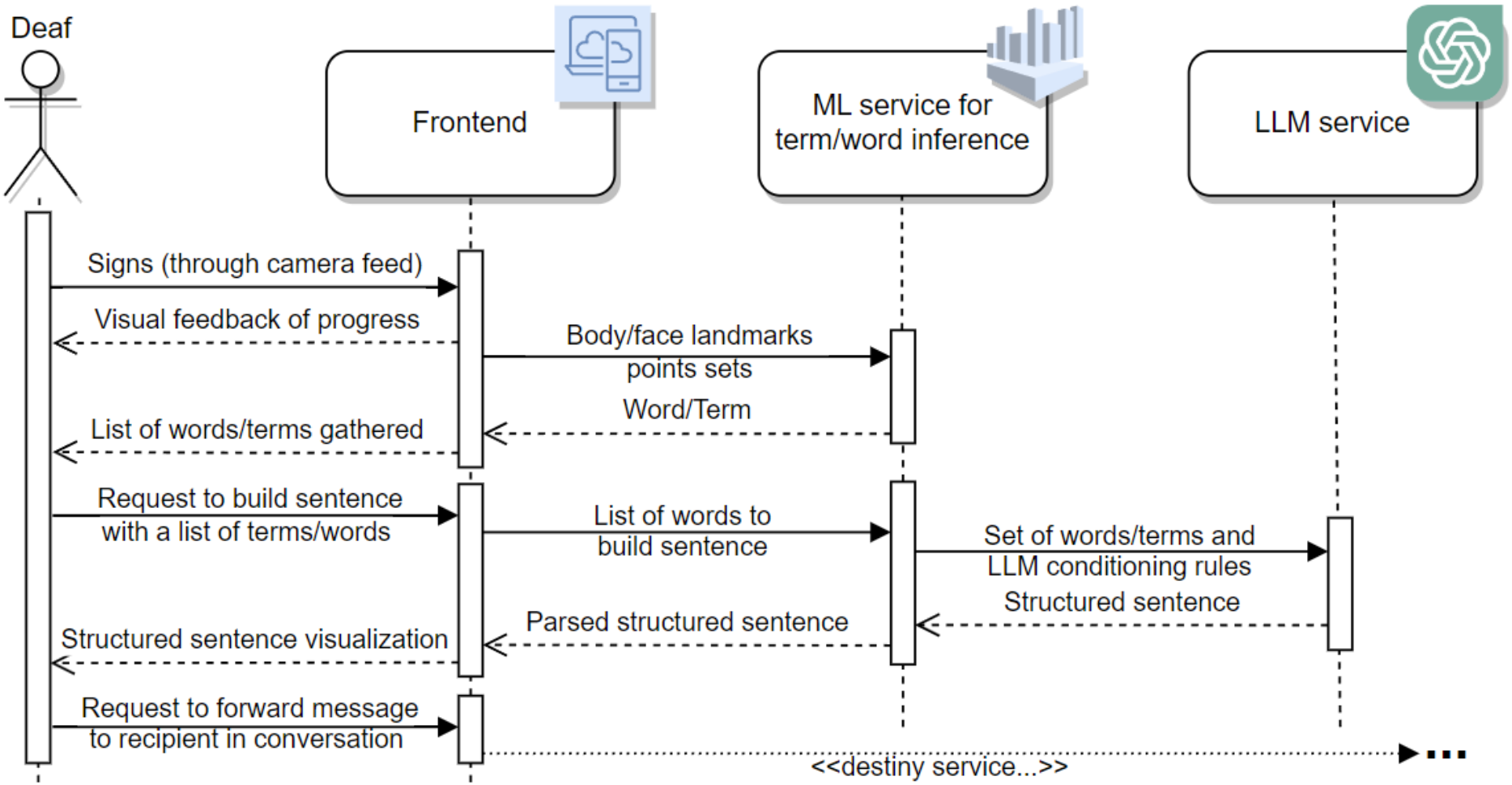
IVLinG: Intérprete Virtual de Língua Gestual

- National project – 3 partners: First solutions, CCG, IPVC
- Aims the creation of a digital platform for virtual and bidirectional interpretation of Portuguese Sign Language (LGP).
- The complete system is composed by two main modules:
  - Real-time virtual interpreter of Portuguese Sign Language (LGP), which will automatically recognize gestures, facial and body expressions, and convert it into text and/or audio. (CCG contribution)
  - Translation of audio into sign language. The response is viewed through an avatar.

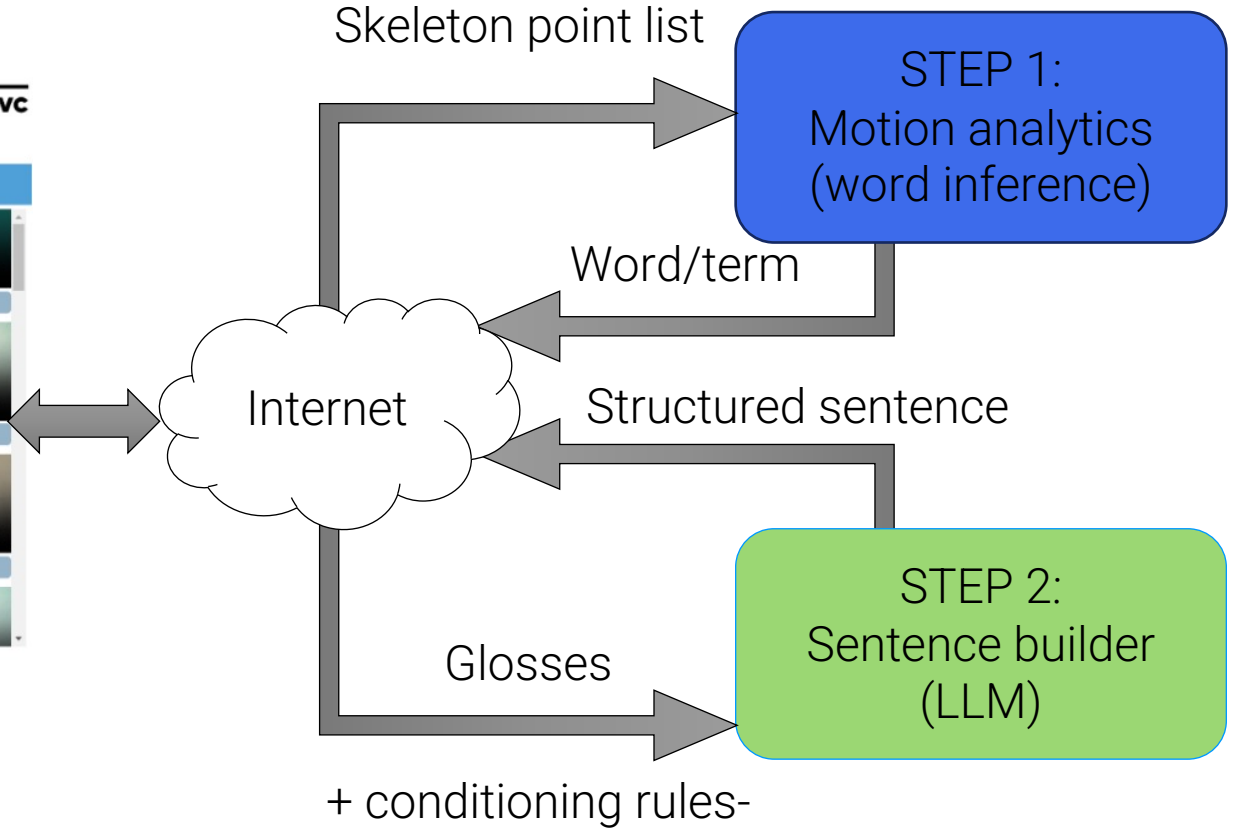
Proposed solution

- A non-invasive Portuguese Sign Language (LGP) interpretation system, as-a-service.
  - Use skeletal posture sequence inference powered by LSTM architectures for sign language interpretation.
  - Use a buffer-based interaction technique to assist in the tokenization of LGP words/terms.
  - Integrate with a large language model (LLM), more specifically, ChatGPT, to transform Deaf loose glossas into structured sentences.

System overview



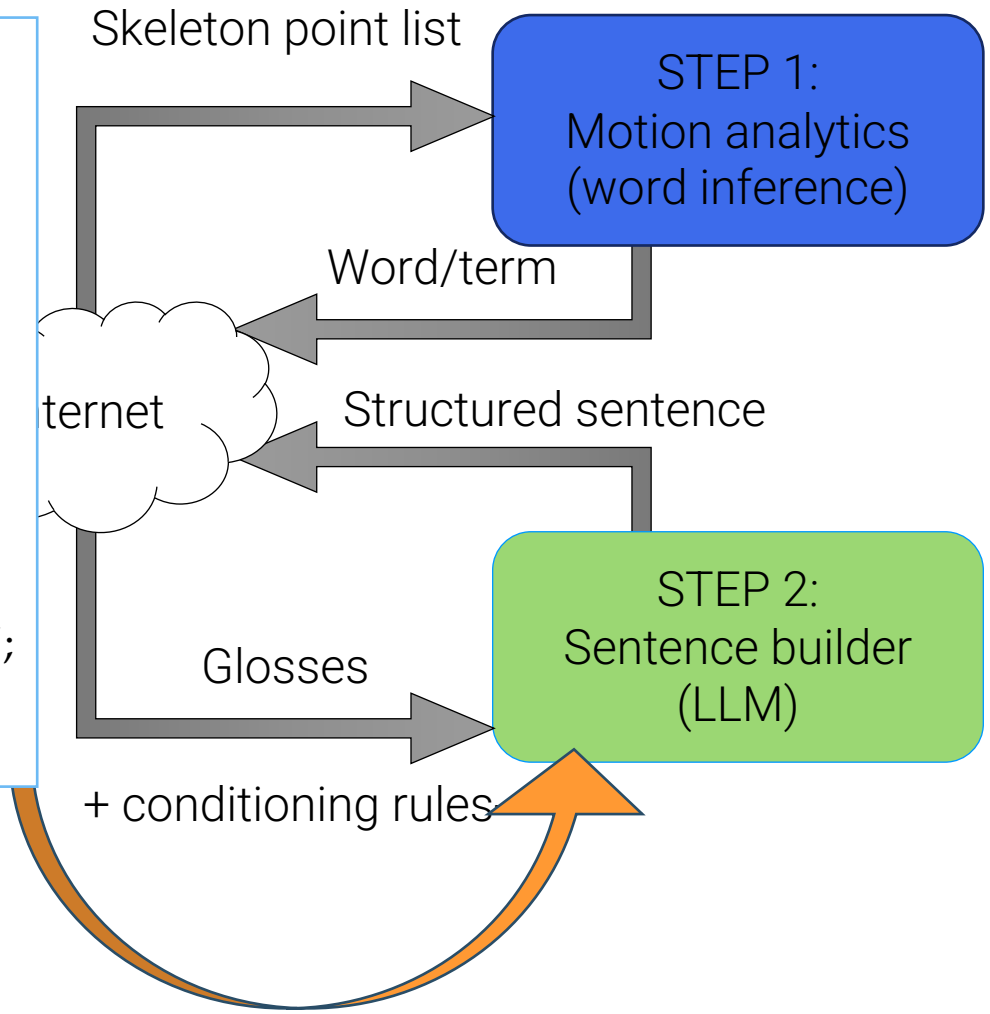
System in practice



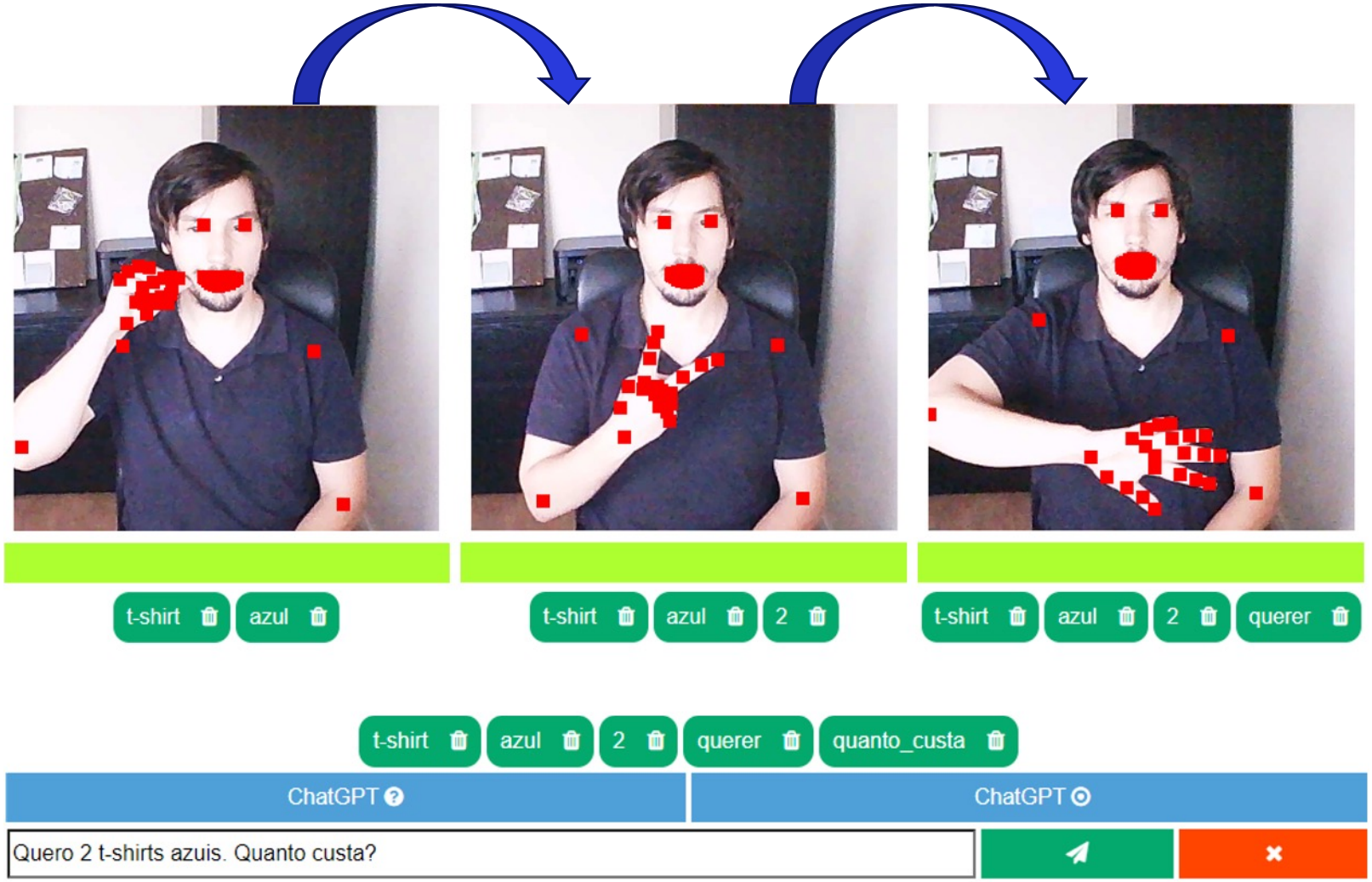
System in practice

Rules:

a) "To ignore repeated words/terms";
b) "To restrain, as much as possible, to the tokens that there are in the given set, avoiding to add more";
c) "To consider the present indicative by default";
d) "If a personal pronoun is not indicated, to conjugate the verbs in the first person singular";
e) "To perform only minimal transformations to the words with the goal of ensuring grammatical correctness";
f) "To perform spelling corrections, whenever necessary";
g) "To correct gender and number agreement inconsistencies";
h) "To interpret numbers as quantifiers".



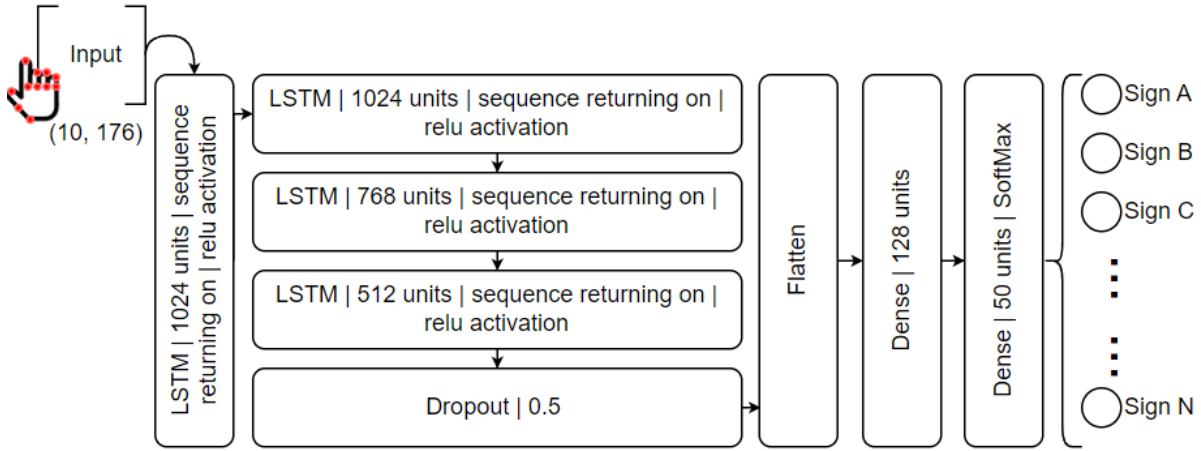
System in practice



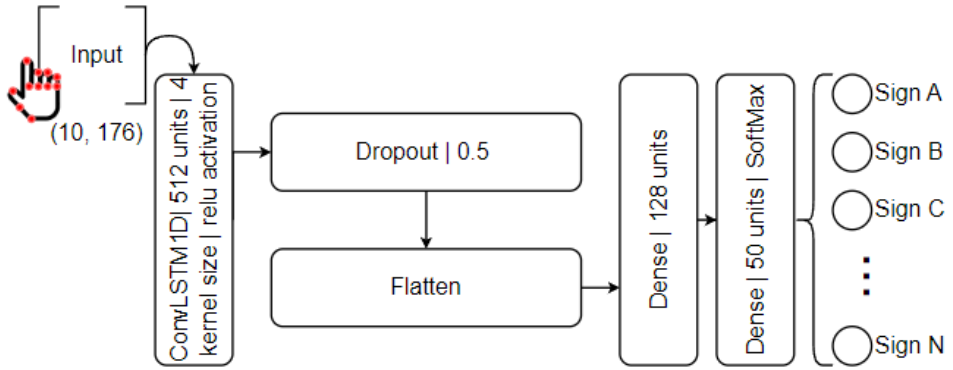
Results

Two LSTM-based structures were used

SimpleLSTM



ConvLSTM



Results

Model	*Augmentation conditions*	*Augmentation limit*	*Accuracy*	*Loss*	*Final Epoch*
SimpleLSTM	*Horizontal flips (HF), not augmented (NA), not balanced.*	*N/A*	*80%*	*1.23*	*155*
	HF + shear operations (SO), not augmented, not balanced.	*N/A*	*87%*	*0.98*	*100*
	HF + SO + RRS augmentation, balanced.	*1008*	***94.8%***	*0.47*	*110*
	HF + SO + SBI augmentation, balanced.	*1008*	*92%*	*0.72*	*100*
	HF + SO + inline RRS/SBI augmentation, balanced.	*1008*	*93.8%*	*0.444*	*75*

Model	*Augmentation conditions*	*Augmentation limit*	*Accuracy*	*Loss*	*Final Epoch*
ConvLSTM	*HF + SO + RRS augmentation, balanced.*	*1008*	***95.6%***	*0.647*	*85*
	HF+ SO + inline RRS/SBI augmentation, balanced.	*1008*	*94.4%*	*0.455*	*150*

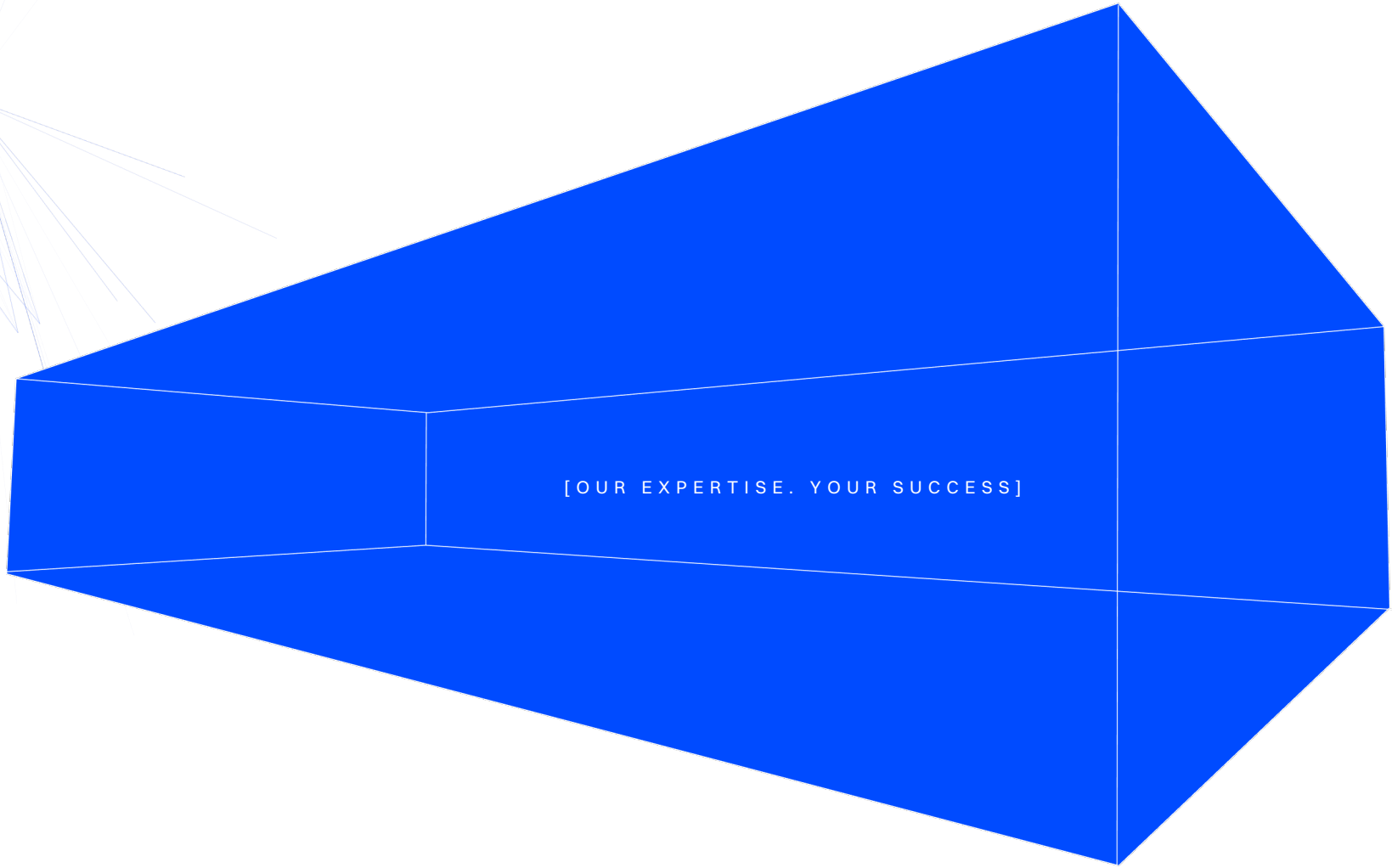
Conclusions

- It was proposed a comprehensive LGP interpretation system, designed as a service
- It was created a LGP dataset, comprising 50 unique terms
- ConvLSTM outperformed SimpleLSTM in terms of accuracy (95.6% vs. 94.8%), but with significant higher training time (17.5 times higher).
- The tests done with ChatGPT were successful, systematically generating text conditioned by prespecified rules with good semantic correlation with expected sentences, around 81% of average matching.

Future work

- Expand the LGP vocabulary beyond the initial 50 terms and, also, increasing the number of examples per sign
- Experimenting with novel augmentation strategies could potentially address scenarios where scarcity of examples persists
- Test with alternative neural network architectures, such as Transformers.
- Refine the conditioning rules of ChatGPT presents an opportunity to enhance the quality of generated sentences.
- Implement an augmented reality frontend (we are currently working on this)

CCG / ZGDV
IT Innovation Centre



+351 253 510 580

info@ccg.pt

UMinho, Campus de Azurém, ed. 14
4800-058 Guimarães . Portugal

www.ccg.pt

PRESIDÊNCIA DO CONSELHO DE MINISTROS
Secretaria-Geral

CTI CENTROS DE TECNOLOGIA E INOVAÇÃO

FCT Fundação para a Ciência e a Tecnologia