

Danae Sanchez Villegas, Ph.D.

📍 Copenhagen, Denmark • 📞 +45 5028 1180 • 📩 danaesavi@gmail.com
LinkedIn: [linkedin.com/in/danae-sanchez-villegas](https://www.linkedin.com/in/danae-sanchez-villegas) • 🌐 danaesavi.github.io • 🎓 [Google Scholar](https://scholar.google.com/citations?user=QWzgkxUAAAAJ&hl=en)

Summary: Applied scientist and postdoctoral researcher specializing in large language models (LLMs) and multimodal AI for text and image data. Experienced in designing, training, and evaluating advanced models and applying them to domains including political communication, medical applications, and e-commerce personalization.

Professional Experience

Research Scientist (Postdoc), Copenhagen University Copenhagen, DK *Feb 2024 – Present*
Applied research on LLMs and multimodal AI for real-world applications (incl. medical diagnostics); designed and evaluated image-to-text architectures; taught graduate-level NLP (transformers, generative models) and supervised MSc projects.

Research Associate, University of Sheffield Sheffield, UK *Aug 2023 – Jan 2024*
Applied multimodal AI and NLP to analyze political ads, integrating text–image data, large-scale dataset creation, and pattern detection in campaign messaging.

Amazon Applied Scientist Intern – Alexa Shopping London, UK *Jun 2022 – Dec 2022*
Developed generative model approaches and LLM prompting strategies to improve customer satisfaction prediction in a production environment.

Data Scientist, Deep Dive Mexico City, MX *Mar 2017 – Sep 2018*
Developed NLP, social media analytics, and geospatial solutions for finance and real estate clients to address business-critical problems.

Education

Ph.D. in Computer Science, University of Sheffield (Sheffield NLP) Sheffield, UK *2019 – 2023*
Thesis: *Beyond Words: Analyzing Social Media with Text and Images*

M.Sc. in Computer Science (Distinction), University of Sheffield Sheffield, UK *2018 – 2019*
Thesis: *Machine Comprehension Using Commonsense Knowledge*

B.Sc. in Computer Engineering (Honorable Mention), ITAM Mexico City, MX *2012 – 2016*
Thesis: *Time Series Modeling for Stock Market Prediction Using Machine Learning Methods*

Skills

- **Programming:** Python, Java, Bash, SQL, Git, Unix/Linux, HPC clusters (Slurm)
- **ML Frameworks:** PyTorch, TensorFlow, Hugging Face Transformers, Keras, scikit-learn, OpenCV, NumPy, Pandas, Matplotlib
- **Model Development:** LLMs, multimodal learning, generative AI, distributed training (Accelerate, FSDP), model evaluation, prompt engineering, dataset curation
- **Applied Domains:** Political communication, medical diagnostics, e-commerce personalization (Amazon Alexa)
- **Languages:** English (Fluent), Spanish (Native), Danish (A2+), German (A1)

Community & Leadership

Speaker: Cohere AI Labs, Gothenburg University, Copenhagen NLP, Mexican NLP Summer School.

Co-organizer: D3A Workshop on Big Scandinavian Data & LLMs, SomosNLP (Iberomerician NLP community), LatinX in AI (global non-profit supporting Latin American researchers in AI).

Selected Publications (full list on Google Scholar)

1. **ImageChain: Advancing Sequential Image-to-Text Reasoning in Multimodal LLMs.** Sanchez Villegas D., Ziegler I., Elliott D. — arXiv:2502.19409, 2025.
2. **Evaluating Multimodal Language Models as Visual Assistants for Visually Impaired Users.** Karamolegkou A., Nikandrou M., Pantazopoulos G., Sanchez Villegas D., et al. — ACL, 2025 **(SAC Award – Human Centered NLP).**
3. **A Multimodal Analysis of Influencer Content on Twitter.** Sanchez Villegas D., Goanta C., Aletras N. — ACL, 2023 **(Area Chair Award – Society & NLP).**
4. **Sheffield’s Submission to the AmericasNLP Shared Task.** Gow-Smith, Sanchez Villegas D. — AmericasNLP, 2023 **(First Place Submission).**