

# LEONARDO V. CASTORINA

✉ leonardo.castorina@outlook.com    📞 REDACTED ONLINE    🌐 universvm    🌐 in/leonardo-castorina

## EDUCATION

### The University of Edinburgh

- Ph.D. in ML for Protein Design (Biomedical AI CDT) Sep 2021 - Present
- MScR. Artificial Intelligence (Distinction) Sep 2020 - Sep 2021
- BSc. (Hons) Biochemistry Sep 2016 - May 2020

## EXPERIENCE

 **NEC Oncolmmunity** – Machine Learning Consultant Nov 2024 - Present

 **Microsoft Research** – Machine Learning Consultant Nov 2023 - Apr 2024

- Disentangled TCR-peptide-MHC interactions to quantify individual contributions in binding.
- Integrated structural and sequence data, discovering multiple disease-associated variants.

 **Microsoft Research** – Research Scientist Intern Jun 2023 - Sep 2023


- Analysed 30K TCR repertoires with MHC and peptide data to identify key interaction patterns.
- Developed statistical method to analyse MHC sequence variance in the context of TCR repertoires.

 **NEC Labs Europe** – Machine Learning Consultant Feb 2023 - Apr 2023

 **NEC Labs Europe** – Research Scientist Intern Oct 2022 - Feb 2023

- Developed GNN models integrating 3D and biological features for TCR-pMHC binding prediction.
- Designed distance-based data splitting algorithm to evaluate model TCR-pMHC generalization.

 **Osmitau Technologies** – CEO & Co-Founder Mar 2019 - Jun 2020

- Led the development of an ML tool for personalised drowsiness detection in cars.
- Managed finances, pitched to manufacturers, co-authored  [white paper](#) on automotive safety.

 **IBM** – Software Engineer Intern Jun 2019 - Sep 2019

- Built an adaptive search radius for vulnerable missing people, reducing search time by ~1.5 hours.

 **P&G** – R&D Intern Jul 2018 - Sep 2018

- Performed market and consumer research to investigate the health benefits of precision vitamins.

## PROJECTS

**TIMED-Design** Deep Learning for de novo Protein Design  GitHub  Try it out

- Designed, benchmarked, and improved 3D CNN models by 10+% for inverse protein folding.
- Implemented user-friendly UI to use models, analyse predictions, and Monte Carlo sampling.
- Winner of the 2023 TensorFlow Community Spotlight Prize out of 1K+ projects.

**Aposteriori** Protein Structures Voxelisation for Deep Learning  GitHub

- Implemented parallel processing, atom encodings, Gaussian representations, and a CLI.

**TEDx** AI in Healthcare: The Next Frontier  YouTube

- Selected to discuss AI and its applications in healthcare. Received 20K views within a year.

**How to Create a Protein** Mini Course on Proteins and Design  GitHub

- Devised open-source course for high school students to teach protein design and 3D print proteins.

## PUBLICATIONS

🔗 TIMED-Design: Flexible and Accessible Protein Sequence Design with CNNs

**Leonardo V. Castorina** 🇮🇹, Suleyman Mert Ünal, Kartic Subr, Christopher W. Wood, (*Protein Engineering and Design*, 2024)

🔗 Assessing the Generalization Capabilities of TCR Binding Predictors via Peptide Distance Analysis

**Leonardo V. Castorina** 🇮🇹, Filippo Grazioli, Pierre Machart, Anja Moesch, Federico Errica, (*bioRxiv*, 2023)

🔗 Attentive Variational Information Bottleneck for TCR-peptide Interaction Prediction

Filippo Grazioli, Pierre Machart, Anja Mösch, Kai Li, **Leonardo V. Castorina** 🇮🇹, Nico Pfeifer, Martin Renqiang Min, (*Bioinformatics*, 2022)

🔗 PDBench: Evaluating Computational Methods for Protein Sequence Design

**Leonardo V. Castorina** 🇮🇹, Rokas Petrenas, Kartic Subr, Christopher W. Wood, (*Bioinformatics*, 2023)

🔗 Deep Attention Super-Resolution of Brain MRI Acquired Under Clinical Protocols

Bryan M. Li, **Leonardo V. Castorina** 🇮🇹, Maria del C. Valdés Hernández, Una Clancy, Stewart J. Wiseman, Eleni Sakka, Amos J. Storkey, Daniela Jaime Garcia, Yajun Cheng, Fergus Doubal, Michael T. Thrippleton, Michael Stringer, Joanna M. Wardlaw, (*Frontiers in Computational Neuroscience*, 2022)

🔗 Metrics for Quality Control of Results From Super-Resolution Machine-Learning Algorithms

**Leonardo V. Castorina** 🇮🇹, Bryan M. Li, Amos Storkey, Maria C. Valdés-Hernández, (*University of Edinburgh. Centre for Clinical Brain Sciences and School of Informatics*, 2021)

🔗 Information Carrier Object and System for Retrieving Information (Patent)

Jake Love, **Leonardo V. Castorina** 🇮🇹, Thomas G. O'Leary, Weilun Gong, Giacomo Giuseppe Chiarella, (*Patent US11132086B1*, 2021)

## BLOG POSTS

**[Towards Data Science] How to Solve the Protein Folding Problem: AlphaFold 2**

**M** Medium

- Awarded Medium Boost prize. Explored AF2 dimensionality and layer functions with visual aids.

**[Towards AI] Latent Diffusion Explained Simply (with Pokémon)**

**M** Medium

- Featured in Towards AI. Used Pokémon to illustrate the concept and applications of diffusion.

**[Better Humans] Obsidian Tutorial for Academic Writing**

**M** Medium

- Nominated Obsidian Gems of the Year (2023). 50K views within a year of publishing.

**[Better Humans] How to Boost Your Productivity for Scientific Research Using Obsidian**

**M** Medium

- Winner of Obsidian Gems of the Year (2022) - Written Content. Received \$400+ and merchandise.
- Featured in the Obsidian Roundup. Received 4K views within one week of publishing.

**[Better Humans] 20+ MacOS Apps to Boost Your Productivity**

**M** Medium

- Named Top 1500 June Article in Medium. Received 4K views within the first month of publishing.

## SKILLS

**Tools:** Python, Git, Jupyter, Streamlit, PyMOL.

**Libraries:** Keras, NumPy, Pandas, PyTorch, SciKit-Learn, TensorFlow.

**Machine Learning:** CNNs, GANs, GNNs, Protein ML/Structural Bioinformatics, Transformers, VAEs.

## LANGUAGES

English (Fluent), Italian (Fluent), French (DELF B2), Sicilian (Fluent).