Manuel Baltieri, Ph.D.
Cybernetics and AI/ALife researcher. Mathematical modeller. Coder.

PROFILE

- Cross-disciplinary researcher developing foundations of intelligent behaviour and agency in cybernetics, AI, ALife, and cognitive science. **Skills**: Bayesian inference, optimal control, ALife, AI/ML.
- Building mathematical models for AI systems that are interpretable and safe for society. **Skills**: applied category theory, geometric deep learning.
- Coding agent-based systems driven by mathematically sound, and safe, definitions and models of decision making. **Skills**: Python (Pytorch/JAX), C, C++, MATLAB, Java, and others.

Tokyo, Japan. **manuelbaltieri.com** manuel.baltieri@gmail.com **github.com/mbaltieri**

EXPERIENCE

Member of Board of Directors Nov 2023 - present
ISAL - International Society for Artificial Life
- **Achievements**: Co-organising 3 international ALIFE conferences, coordinating international workshop on "Models of consciousness" with ALIFE conference, making ALife a main topic of AI safety in Japan,

Chief Researcher (previously Researcher) Nov 2021 - present
Araya Inc., JP
- **Achievements - 1**: NeuroAI framework and models of safe human driving for world-leading car manufacturer. The models steered the client's experimental designs, resulting in a >6 month large-scale revision plan with future applications to autonomous vehicle systems,
- **Achievements - 2**: Writing (4) papers on mathematical framework for interpretable AI (applied category theory, optimal control, Bayesian inference and reinforcement learning) + podcasts (2) + invitations (4) to workshops on AI safety and societal impact,
- **Achievements - 3**: Leading client-facing consulting contract (+ renewal of major contract), development of custom LLM solutions in Pytorch (architecture optimisation, fine-tuning, prompt engineering and performance metrics selection) reducing parameter usage by 20% while maintaining performances.

Research Consultant and Developer Sep 2021 - Nov 2021
Nested Minds Network Ltd., USA
- **Achievements**: First Numpyro implementation of active inference for continuous-state continuous-time state-space models (with > 10 projects built on it), leading to transition from Pytorch/Numpy to JAX.

JSPS/Royal Society Postdoctoral Research Fellow Nov 2019 - Oct 2021
Laboratory for Neural Computation at the Center for Brain Science, RIKEN, JP
- **Achievements**: Developeds *'Generalised Filtering'*, a package for continuous-state continuous-time state-space models in Pytorch for Bayesian inference, learning and optimal control, used as a basis for active inference at Nested Minds Netework. Published 8 papers in AI/ML/ALife, control theory and neuroscience, gave > 5 presentations.

Honorary Senior Visiting Researcher Nov 2019 - present
University of Sussex, UK

Postdoctoral Research Fellow Mar 2019 - Oct 2019
University of Sussex, UK
- **Achievements**: First implementation of active inference (via Bayesian neural networks in Python) in wheeled robots (ROS) + master student thesis supervision (psychophysics).

TEACHING

Lecturer in Mathematical Foundations of Agency Dec 2022 - Feb 2023
CHAIN - Hokkaido University, JP
Teaching mathematical approaches to model agency for the CHAIN Winter Graduate School.

Lecturer in Machine Learning Feb 2019 - Jun 2019
University of Sussex, UK
Teaching undergraduate courses in Neural Networks and Fundamentals of Machine Learning.

Teaching Assistant 2015 - 2019
University of Sussex, UK
Computer Vision, Artificial life, Further Programming (Java), Mathematical concepts.

Education

Ph.D. in Informatics and Artificial Intelligence, University of Sussex, UK 2019

- **Achievements**: Mathematical analysis and Python implementation of NeuroAI-based free energy principle/active inference framework with applications to optimal control theory. Published 8 papers, gave > 10 presentations (2 invited keynotes for active inference workshops).

M.Sc. in Evolutionary and Adaptive Systems, University of Sussex, UK 2014

- **Achievements**: Mathematical analysis of free energy principle and implementation in C/C++.

B.Eng. in Information and Business Organisation, Università di Trento, IT 2012

- **Achievements**: Implementation of extension of Tabu Search for portfolio optimisation in C/C++.

Service

Guest Editor for Special Issue on ALIFE 2023 + Proceedings of ALIFE 2023 Conference + ALife Journal and Entropy - MDPI

Programme Organiser of ALIFE 2025 (Kyoto, JP), 2025
Co-Organiser of *Goal-Directed behavior in life and non-life* at ALIFE 2024 (Copenhagen, DK), 2024
Research Organiser of *Mathematical Boundaries Workshop* (Berkeley, USA), 2024
Proceedings/Local Chair of ALIFE 2023 (Sapporo, JP), 2023
Achievements: Proceedings published with MIT Press.
Keynote speakers Chair of ALIFE 2022 (online), 2022
Lead Co-Organiser of *"Hybrid life: Approaches to integrate biological, artificial and cognitive systems"*, 2018-2022
Achievements: Created and managed five special sessions, speakers' invitations, reviewing and editing processes of submissions published in "Hybrid Life: Integrating Biological, Artificial, and Cognitive Systems".
Co-Organiser of *"ALife for Social and Environmental Good"* at ALIFE 2020 (online), 2020

Academic reviewer for tens of journals + conferences (highlights: Nature Communications, PNAS, ALIFE, IROS, ICLR (Workshop))

Awards and grants

Fellowships and scholarships

- JSPS/Royal Society Postdoctoral Fellowship (**USD 80,000/ 2 years** from JSPS, JP, selected by Royal Society, UK) 2019 - 2021
- Postdoctoral Fellowship (**USD 140,000/ 3 years** from oLife (Origins of Life) Fellowship Programme, The Netherlands; ***withdrawn*** *to accept JSPS* ~~2019-2022~~
- EON-ELSI Visiting Graduate Researcher Scholarship (**USD 5,000/ 2 months** from ELSI, Tokyo, JP) Dec 2017-Jan 2018
- PhD Scholarship (**USD 60,000/ 3 years** from University of Sussex, UK) 2015-2018

Grants

- ARIA (Co-PI) - Mathematics for Safe AI (**USD 380,000**) 2025 - 2027
- Kakenhi (PI) - Grant-in-Aid for Scientific Research (**USD 22,000**) 2019 - 2021

Journal publications (* Corresp. author)

- Torresan, F. & **Baltieri, M.*** (2024). Disentangled representations for causal cognition. *Physics of Life Reviews*.
- **Baltieri, M.***, Iizuka, H., Witkowski, O., Sinapayen, L., & Suzuki, K. (2023). Hybrid Life: Integrating Biological, Artificial, and Cognitive Systems. *WIREs Cognitive Science*. **(Invited)**
- Bruineberg, J., Dolega, K., Dewhurst, J. & **Baltieri, M.*** (2022). The Emperor is Naked: Replies to the commentaries on the target article. *Behavioral and Brain Sciences*.
- Bruineberg, J., Dolega, K., Dewhurst, J. & **Baltieri, M.*** (2021). The Emperor's New Markov Blankets. *Behavioral and Brain Sciences*.
- Mannella, F., Maggiore, F., **Baltieri, M.** & Pezzulo, G.* (2021). Active inference through whiskers. *Neural Networks*.
- Hipólito, I., **Baltieri, M.**, Friston, K., & Ramstead, M. J.* (2021). Embodied skillful performance: Where the action is. *Synthese*, 1-25.
- **Baltieri, M.***, & Buckley, C. L. (2019). PID control as a process of active inference with linear generative models. *Entropy, 21(3)*, 257.

Preprints and Work in progress (* Corresp. author)

- **Baltieri, M.*** & Rosas, F. (Under review). What is it like to be a Braitenberg vehicle?
- Virgo, N.*, **Baltieri, M.**, Biehl, M. & Capucci, M. (Under review). A "good regulator theorem" for embodied agents.

- **Baltieri, M.***, Biehl, M., Capucci, M. & Virgo, N. (2025). A Bayesian interpretation of the internal model principle.
- **Baltieri, M.*** & Isomura, T. (2021). Kalman filters follow the natural gradient of free energy at steady-state. *arXiv pre-print arXiv:2111.10530*
- **Baltieri, M.*** & Buckley, C. L. (2020). On Kalman-Bucy filters, linear quadratic control and active inference. *arXiv pre-print arXiv:2005.06269*
- McGregor, S., **Baltieri, M.** & Buckley, C. L. (2015). A minimal active inference agent. *arXiv pre-print arXiv:1503.04187*

Conference proceedings (* Corresp. author)

- Rosas, F., Boyd, A., & **Baltieri, M.** (2025). AI in a vat: Fundamental limits of efficient world modelling for agent sandboxing and interpretability. In *Reinforcement Learning Conference - RLC 2025*.
- **Baltieri, M.***, Buckley, C. L., & Bruineberg, J. (2020). Predictions in the eye of the beholder: an active inference account of Watt governors. In *Artificial Life Conference Proceedings - ALIFE 2020*. MIT Press.
- **Baltieri, M.*** (2020). A Bayesian perspective on classical control. In *2020 International Joint Conference on Neural Networks - IJCNN 2020*. IEEE.
- Tschantz, A., **Baltieri, M.**, Seth, A. & Buckley, C. L. (2020). Scaling Active Inference. In *2020 International Joint Conference on Neural Networks - IJCNN 2020*. IEEE.
- **Baltieri, M.***, & Buckley, C. L. (2019). Active inference: computational models of motor control without efference copy. In 2019 Conference on Cognitive Computational Neuroscience.
- **Baltieri, M.***, & Buckley, C. L. (2019). The dark room problem in predictive processing and active inference, a legacy of cognitivism? In *Artificial Life Conference Proceedings - ALIFE 2019*. MIT Press.
- **Baltieri, M.***, & Buckley, C. L. (2019). Nonmodular architectures of cognitive systems based on active inference. In *2019 International Joint Conference on Neural Networks - IJCNN 2019*. IEEE.
- **Baltieri, M.***, & Buckley, C. L. (2018). A probabilistic interpretation of PID control. In *International Conference on Simulation of Adaptive Behavior - SAB 2018)*. Springer.
- **Baltieri, M.***, & Buckley, C. L. (2018). The modularity of action and perception revisited using control theory and active inference. In *Artificial Life Conference Proceedings - ALIFE 2018*. MIT Press.
- **Baltieri, M.***, & Buckley, C. L. (2017). An active inference implementation of phototaxis. In *European Conference on Artificial Life Proceedings - ECAL 2017*. MIT Press.

Invited talks

A structural view of predictive coding, Development and validation of a unified theory of prediction and action, Institute of Science, Tokyo, JP — May 2025

Life and its relatives, Breakthrough Initiatives, Oxford, UK — Apr 2025

When does a system model another system?, Internet of Brains Event, Tokyo, JP — Mar 2025

What do Braitenberg vehicles believe?, Computational Agency workshop, CHAIN, Hokkaido University, Sapporo, JP — Jan 2025

Using goals to compress models, effects of bisimulations for (PO)MDPs on embodied agents, Theoretical and Experimental Approaches to Goal-Directed Behaviour, BCAM, Bilbao, SP — Oct 2024

Foundations of AI: building reliable, trustworthy and agentic systems, FUTURA @ Italian Chamber of Commerce in Japan, Tokyo, JP — Oct 2024

The Role of the Free Energy Principle in AI Safety: Markov Blankets and Beyond, TAIS - Technical AI Safety Conference, Tokyo, JP — Apr 2024

LLMs beyond Transformers (panel discussion), Tokyo AI Talks @ GHOVC, Tokyo, JP — Jan 2024

The Emperor's New Markov Blankets, COGS - University of Sussex, Brighton, UK — Oct 2023

Inference with and within a model, CHAIN Seminars, Hokkaido University, Sapporo, JP — Feb 2023

Lecture series - Defining agency, A history of studies of agency, First-principles definitions of agents, CHAIN Winter School on Minimal Cognition and Agency, Hokkaido University, Sapporo, JP — Feb 2023

Inference with and within a model, Consciousness Club, Tokyo, JP — Jul 2022

Variational inference in agents, with connections to control theory and cognitive (neuro)science, RIKEN AIP, Tokyo, JP — Jul 2021

Agency in 100 years, a solved problem?, Discussion with Kevin O'Regan at Tokyo ALIFE 2020 (University of Tokyo and Google Brain), Tokyo, JP (cancelled due to COVID-19) Mar 2020

Active inference for cognitive science and artificial intelligence - open questions and new challenges, CHAIN Academic Seminars, Hokkaido University, Sapporo, JP Jan 2020

PID control as active inference: what can we gain from this formulation?, 2nd Active inference workshop, TU Delft, Delft, NL Nov 2018

The free energy principle and active inference, connecting control theory to biology, Active inference workshop, TU Delft, Delft, NL Apr 2018