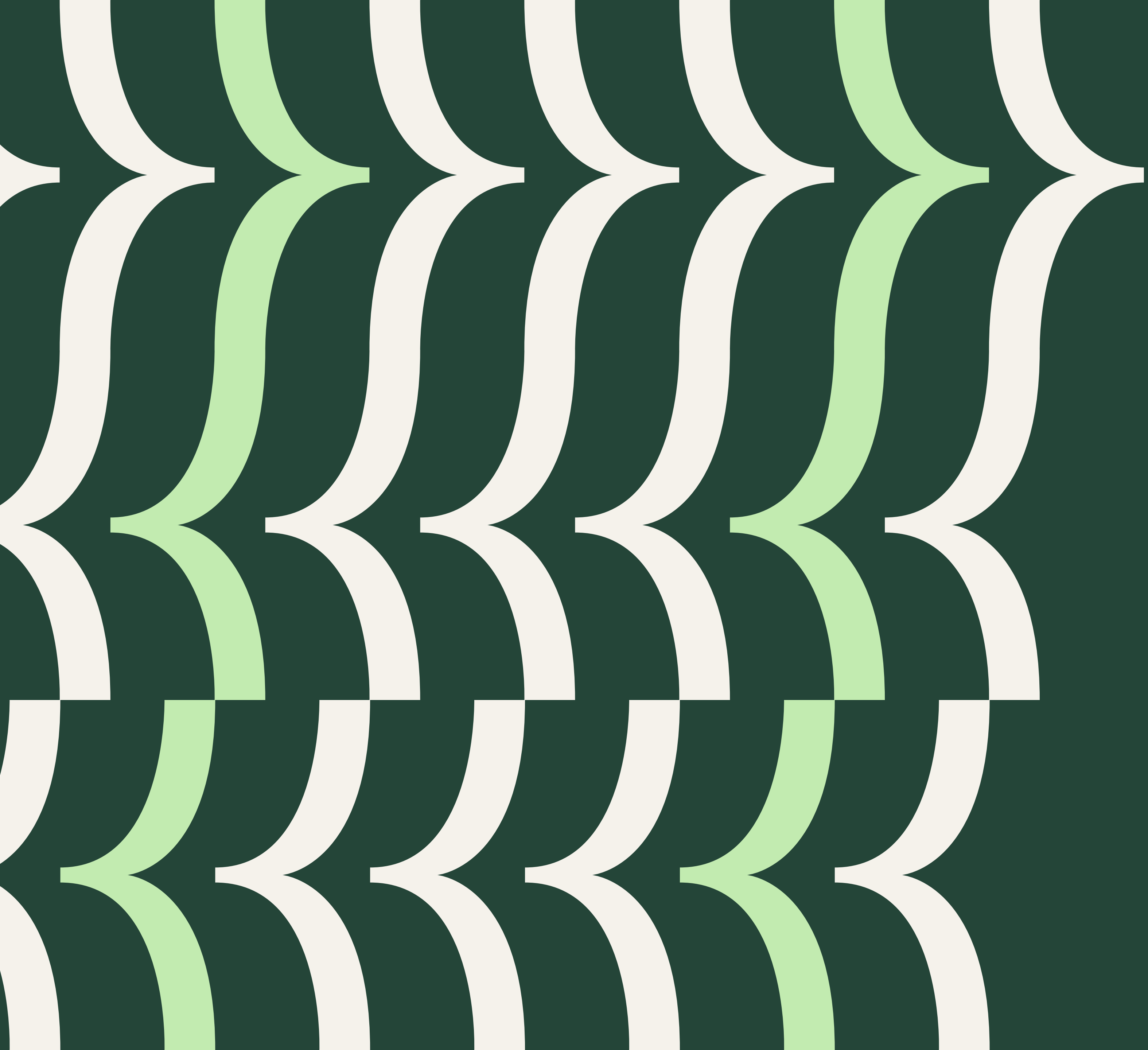




Annual Report 2024

integreat

The Norwegian centre for
knowledge-driven machine learning

Content

About Integreat	4	Events	41	Seminars	58
Organisation	5	Meeting the public	47	The way ahead	59
Research organisation	12	Researcher training programme	53	Integreat voices	60
Research highlights	14	DEI	56	Photowall	62
Publications	35			Integreat in figures and numbers	67

Here we are



A handwritten signature in dark ink, appearing to read 'Arnaldo Frigessi'.

Integreat Centre Director
Arnaldo Frigessi,
Oslo, March 2025

Unlike purely data-centric machine learning, Integreat develops theories, methods, models, and algorithms that integrate general and domain-specific knowledge with data, laying the foundations of knowledge-driven machine learning.

This will be done by combining mathematical and computational cultures, and the methodologies of machine learning, statistics, logic, language technologies, ethics, and mathematics.

Integreat will solve fundamental problems in science, technology, health, and society, and contribute to making Norway a world-leading country in AI.

Welcome to the first annual report of Integreat, the Norwegian Centre for Knowledge-driven Machine Learning! You will find a summary of our first one-and-a-half year, the time of building, bonding, sowing, and strategising; starting from our original plans, influenced by the all-changing developments in generative artificial intelligence (AI), and motivated by our learning across disciplines. The Integreat we have today is an ambitious, solid, well-functioning, forward-looking, exciting and *hyggelig* research centre. We are about 100 people, from our first master's students, our fantastic cohort of doctoral

research fellows, funded by Integreat and other sources, postdocs, researchers, associate and senior professors, to our superb administrative staff - all working with passion and engagement, creativity and originality, competence, and technical strength. If you walk into our hubs in Oslo and Tromsø in Norway, you will see and feel the intellectual, sensory, sentimental, and physical energy. It is turning central open problems in machine learning (ML) into new forms to be solved with new instruments, producing results we have not seen earlier. We will contribute to making ML more precise, considering its uncertainty, and making it more sustainable, fairer, and more explainable. This is needed to develop AI in the right direction, where it is an instrument for fair, just and shared development, for democracy – for a better world for all. This becomes increasingly important in times when AI concentrates in the hands of powerful and venture-capital-driven forces.

Knowledge-driven ML is an active area of research worldwide, and we are well-positioned to contribute significantly, due to our multi-disciplinary approach, with statistics, logic, language technology, theoretical computer science, mathematics and of course ML itself. In these first years of collaboration, we have begun with building powerful approaches that interweave inductive ML, which learns about the world from examples; deductive logic, with rules that describe knowledge, together with logical reasoning; and inferential statistics, which learns relations and enables reasoning about uncertainty. Our science is diverse, with theorems and experiments, theoretical guarantees and benchmark and real-world validation, starting from domain problems or arriving there, publishing in journals, books and conferences, relying on Graphics Processing Units (GPUs) for intensive computing and High-Performance Computing (HPC) systems. The way we understand knowledge is changing, including knowledge about the future use of algorithms,

their future environments and the future data, knowledge about the chain-of-thought principles, knowledge carried by mathematical properties, knowledge about neglected sources of uncertainty, and also false knowledge.

Open AI is fundamental. The international community of AI workers share code and data and participate in open shared challenges that allow to compare methods. This openness is our strength and is essential to develop AI transparently and democratically. We see very powerful resistance to openness, expressing fear that openness is progressive. In fact, it is. Research at Integreat is open and it contributes to the shared development of AI.

An SFF Centre of Excellence provides a fantastic opportunity to tackle challenges that require more time and collaboration to solve. It allows us to work in parallel on many projects that gain from reciprocal vicinity, with unexpected synergies breaking out. Because of increased and recognised researcher-centred autonomy, a SFF can launch innovative strategies, like ways to organise the PhD community, to relate to public and private sectors, in the work with the media and general public, and in the organisation of research. Integreat is experimenting in all these directions, with enthusiasm and signs of success. And of course, as a SFF, we experience recognition, which sometimes allows us to bring our ideas to important discussions. For example, “Stand Up for Science” campaign.

In the upcoming year, we are full of expectations, contributing with results that will make a significant impact on ML, and make us proud, with our PhD fellows blooming, and with increased international visibility. Stay tuned, and please contact me. I look forward to meeting you and finding interesting points of contact.

About Integreat

Integreat is the Norwegian research centre on knowledge-driven machine learning (ML), funded by the Research Council of Norway (2023–2033) under the SFF scheme of centres of excellence. The centre is hosted by the University of Oslo, with a hub in Oslo with researchers from the University of Oslo, the Norwegian Computing Centre (NR) and the Oslo University Hospital, and a hub in Tromsø at UiT-The Arctic University of Norway.

Integreat transforms machine learning by integrating general and domain-specific

knowledge into learning, moving beyond conventional data-centric approaches. By combining statistics, language technology, logic, mathematics, ethics, and machine learning, Integreat develops new theories, models, and algorithms that make ML more accurate, explainable and trustworthy, sustainable, and it quantifies uncertainty better. Our research has broad applications in science, technology, health, and society. With top researchers and strong international collaborations, Integreat contributes to shaping the future of ML.

Integreat will solve fundamental problems in science, technology, health, and society, and contribute to make Norway a world-leading country in AI.

– Integreat Centre Director Arnoldo Frigessi

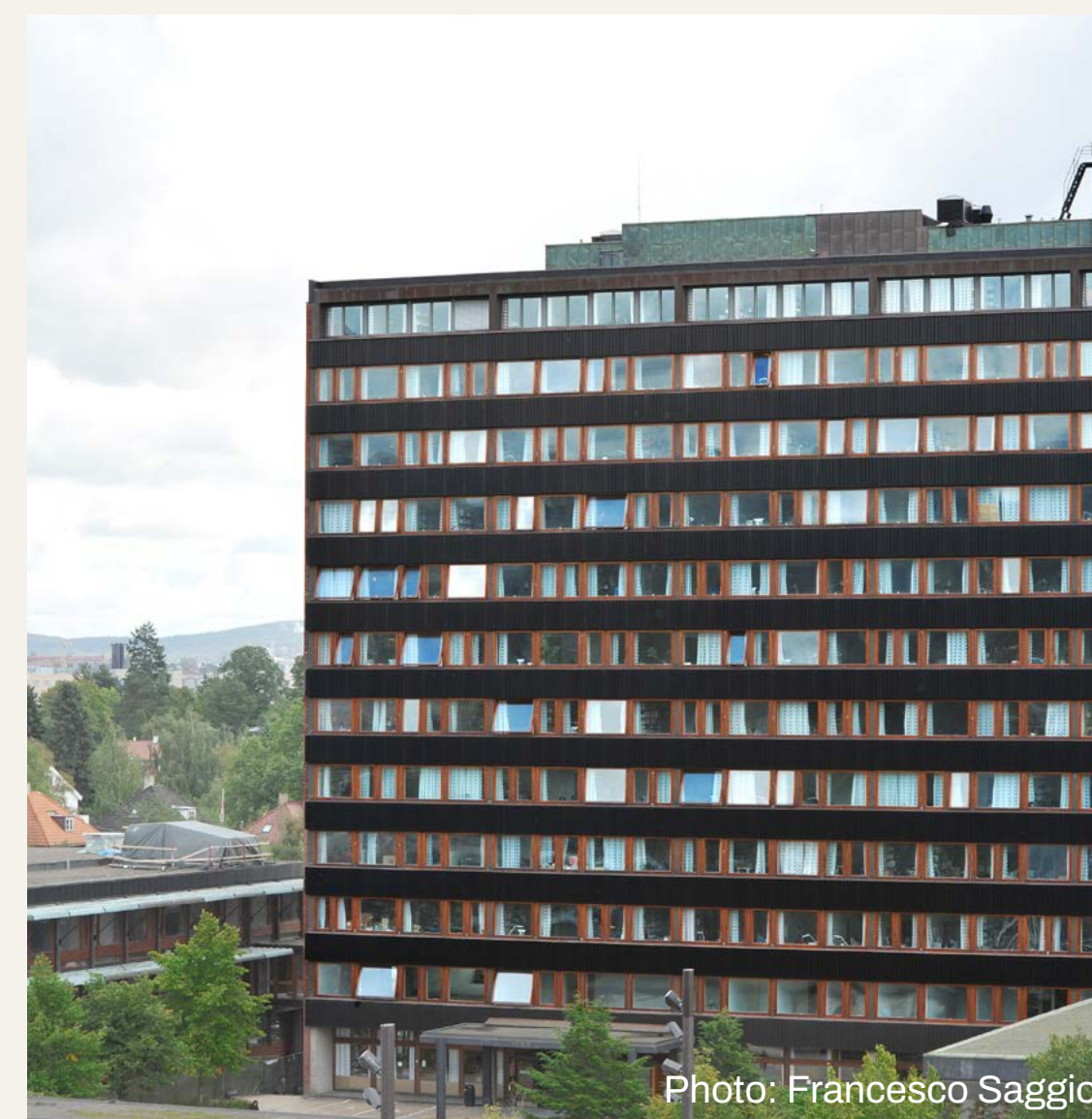


Photo: Francesco Saggio

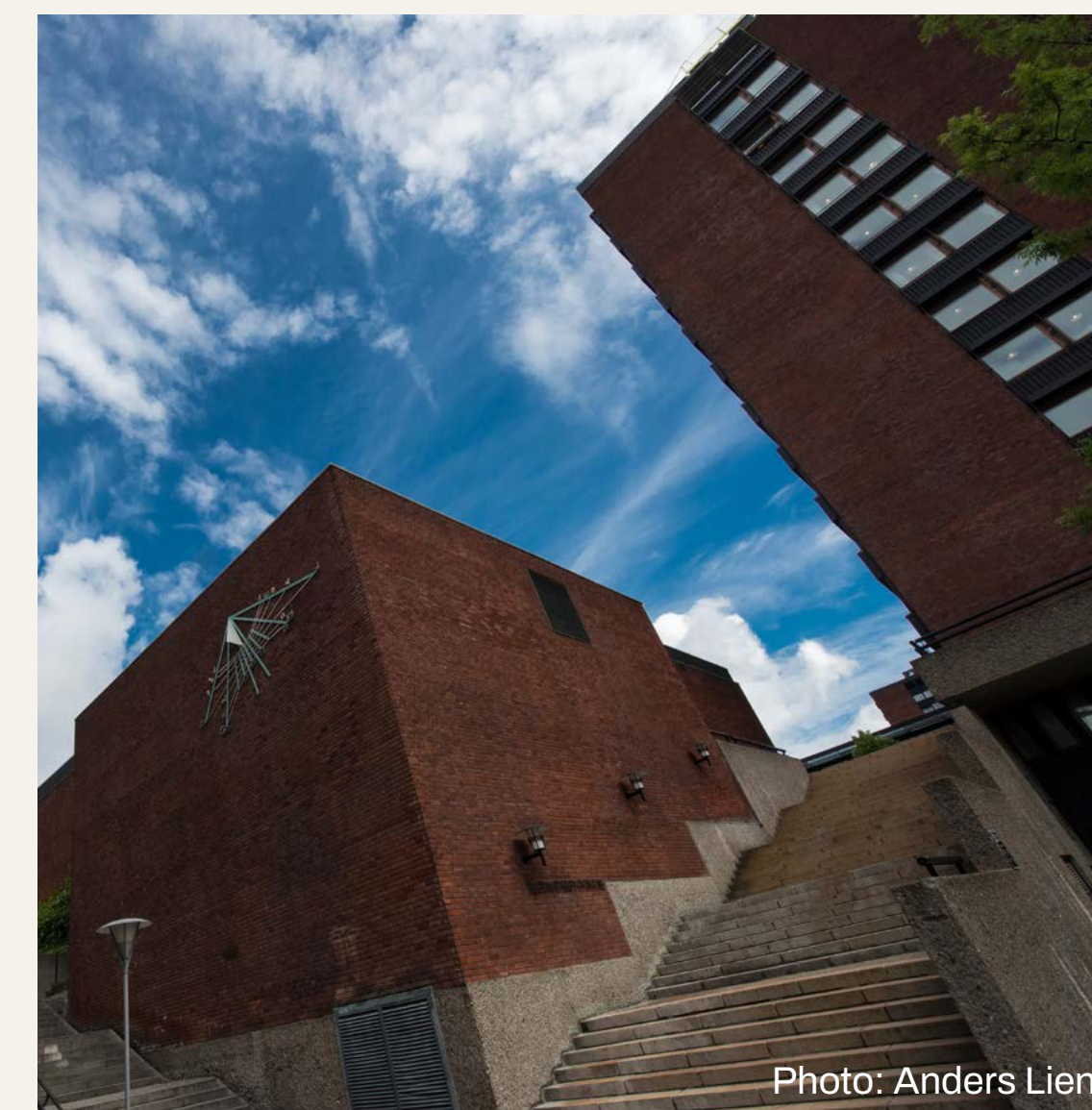


Photo: Anders Lien



Integreat premises University of Oslo, Blindern campus, Niels Henrik Abels building

Organisation

The centre was established in September 2023. At the beginning of 2025, the centre consists of approximately 100 people, including researchers ranging from master's students to professors, as well as a group of administrative staff. At the University of Oslo, researchers are affiliated with the Faculty of Medicine (Institute of Basic Medical Sciences and the Oslo Centre for Biostatistics and Epidemiology), the Faculty of Mathematics and Natural Sciences (Department of Mathematics and Department of Informatics), and the Faculty of Humanities (Department of Philosophy, Classics, History of Art and Ideas). At UiT-The Arctic University of Norway, researchers are based in the Department of Physics and Technology within the Faculty of Science and Technology. A group of researchers is also affiliated with The Norwegian Computing Centre, which is the third key partner of Integreat.

Management

Integreat is led by a group of co-directors:

- **Professor Arnoldo Frigessi**, Director, Oslo Centre for Biostatistics and Epidemiology (OCBE) and Department of Mathematics, University of Oslo
- **Professor Ingrid Glad**, Department of Mathematics, University of Oslo
- **Professor Lilja Øvrelid**, Department of Informatics, University of Oslo
- **Professor Martin Giese**, Department of Informatics, University of Oslo
- **Professor Robert Jenssen**, Machine Learning Group, UiT-The Arctic University of Norway, Tromsø

This group functions as the executive board of the centre. Together with the administrative leader, Maria Dikova, the management team ensures the effective operation and strategic direction of the centre. **The board of Integreat** is chaired by the dean of the Faculty of Mathematics and Natural Sciences of UiO. The deputy chair is the dean of the Faculty of Natural Sciences and Technology of UiT. Other members of the board include the CEO of the Norwegian Computing Centre (NR), the head of the Department of Mathematics, the head of Department of Informatics, and the head of the Institute of Basic Medical Sciences.



Integreat co-directors at the annual retreat in Tromsø in 2024.
From the left Robert Jenssen, Arnoldo Frigessi, Ingrid K. Glad, Martin Giese, and Lilja Øvrelid.

Integreat Board 2024

CHAIR Bjørn Jamtveit Vice Dean of Research, <i>the Faculty of Mathematics and Natural Sciences, University of Oslo</i>	DEPUTY CHAIR Olav Gaute Hellesø Head of <i>Department of Physics and Technology, UiT -The Arctic University of Norway</i>	BOARD MEMBER André Teigland CEO, <i>Norwegian Computing Center (NR)</i>	BOARD MEMBER Geir Dahl Head of <i>Department of Mathematics, University of Oslo</i>	BOARD MEMBER Lene Frost Andersen Head of <i>Department, Institute of Basic Medical Sciences, University of Oslo</i>	BOARD MEMBER Stephan Oepen Head of <i>Department of Informatics, University of Oslo</i>
--	---	---	---	---	---



Integreat administration at the annual retreat 2024.
From the left Olesia Fesenko, Christoffer Hals, Inger Solheim, Maria Dikova, and Petter Bjørklund.

Administrative and technical support

Integreat is supported by administrative and technical staff across all partnering institutions, assisting with procurement, finance, HR, and communication. In 2024, the core administrative team together with

Maria Dikova, included Olesia Fesenko, Øyvind Mørkedal and Christoffer Hals at the University of Oslo, as well as Inger Solheim and Petter Bjørklund at UiT.

Research organisation

Integreat's research is organised into nine interconnected Research Themes (RTs), each addressing specific methods and challenges in knowledge-driven machine learning. These themes align with the centre's four overarching objectives, ensuring a comprehensive and interdisciplinary approach to advancing the field.

See page 12.

Each RT is guided by a team of Principal Investigators (PIs) who steer scientific direction and oversee research progress. Integreat's research follows a project-based approach, with each RT encompassing projects that align with its overarching mission. Additionally, projects are connected to multiple RTs.

« *The cross-disciplinary collaboration will hopefully make us come up with novel answers, but it has already certainly made us ask new kinds of questions.* »

– PI Professor Erik Velldal, Department of Informatics, UiO

Scientific Advisory Committee

The Scientific Advisory Committee (SAC) plays a crucial role in advising on research strategy that leads to scientific excellence. Composed of leading international experts in machine learning, statistics, logic, and language technology, the SAC provides independent advice on the centre's research directions, methodologies, international visibility and impact, providing perspectives on emerging trends and challenges in ML. The SAC helps us to evaluate our research results. In 2024, the SAC held two meetings.

Scientific Advisory Committee:

- **Professor Sofia Olhede**, EPFL Lausanne, Switzerland
- **Professor Barbara Plank**, Ludwig Maximilian University of Munich, Germany
- **Professor Ian Horrocks**, University of Oxford, UK
- **Professor Sach Mukherjee**, University of Cambridge, UK and Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Bonn, Germany
- **Professor Christopher Nemeth**, Lancaster University, UK



[Visit webpage](#)



SAC visiting Integreat in Oslo, 4-5 March 2024. From the left Arnoldo Frigessi, Sofia Olhede, Sach Mukherjee, Christopher Nemeth, Robert Jenssen, Martin Giese, Ingrid Glad, and Barbara Plank.

« *Integreat is a pioneering Centre of Excellence in knowledge-driven machine learning that fuses rigorous statistical, logical, and computational methodologies to tackle fundamental challenges across science, technology, health, and society. Integreat has a strong, multidisciplinary research team with a steadfast commitment to sustainable, ethical, and trustworthy AI, which together position Norway at the forefront of global artificial intelligence innovation.* »

– Integreat SAC member Christopher Nemeth, Professor in Probabilistic Machine Learning, Lancaster University.

Integreat young researchers' forum, INTREF

The Integreat young researchers' forum (INTREF) is the organisation of the PhD and postdoctoral fellows at Integreat. It is led by the elected INTREF Fellows Committee, which represents the interests of its members. See page 56 for a presentation of INTREF. The committee consists of four representatives, and is supported by the INTREF coordinator and by the Integreat Administrative leader.

This is the current composition:

INTREF Fellows Committee:

- **Ingrid Hobæk Haff**,
PI, Department of Mathematics, UiO
- **Clara Bertinelli Salucci**,
Postdoctoral Fellow,
Department of Mathematics, UiO
- **Emil Poiesz**, PhD Fellow,
Department of Informatics, UiO
- **Johan Mylius-Kroken**, PhD Fellow, UiT
- **Kushagri Tandon**, PhD Fellow,
Department of Mathematics, UiO

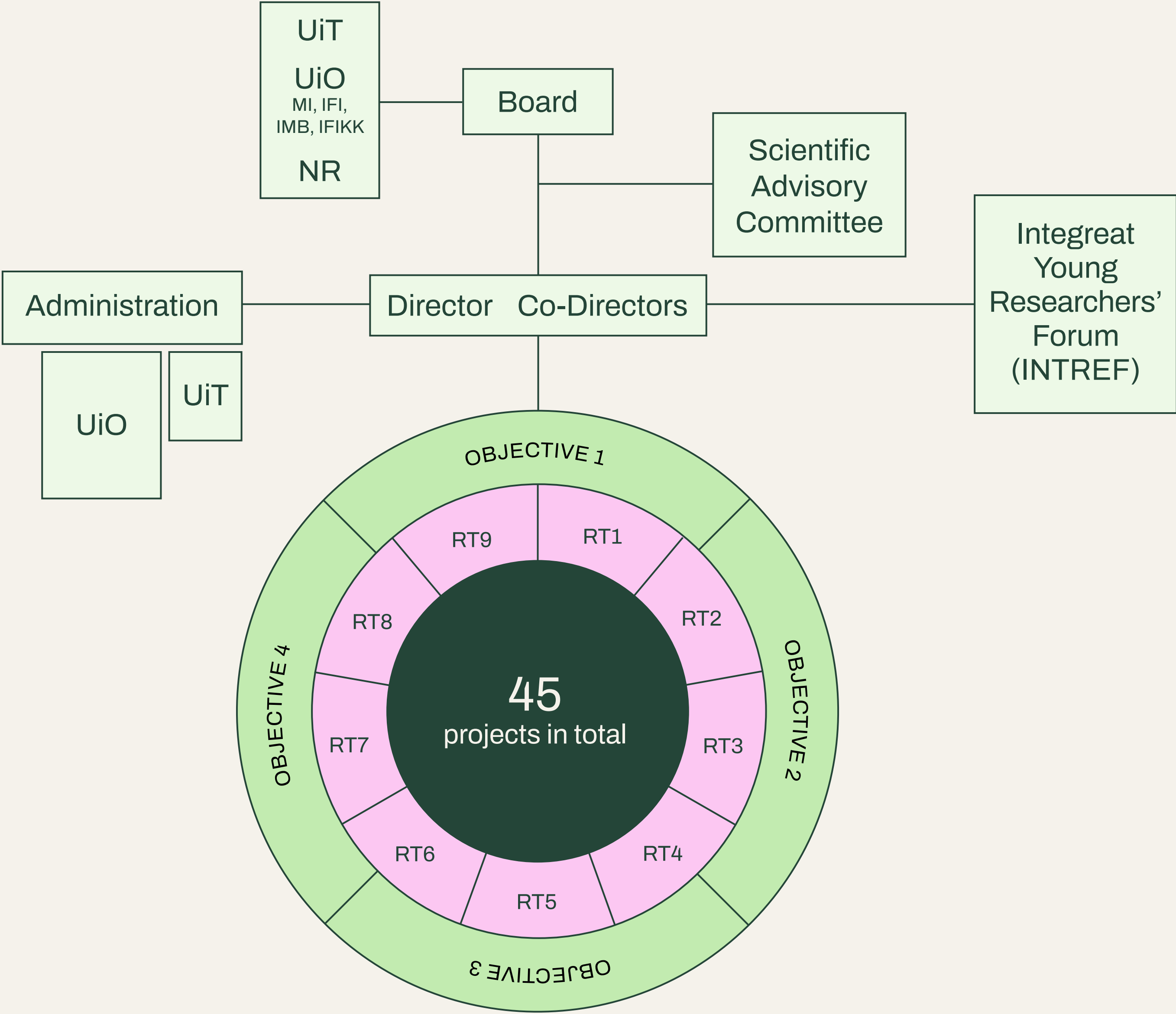


INTREF Fellows committee meeting the coordinator Ingrid Hobæk Haff and the director Arnaldo Frigessi.

Diversity, Equity and Inclusion committee

This committee consists of the Co-Directors, Integreat Administrative leader, and Integreat's inclusion officer PhD Fellow Elisabeth Maria Griesbauer. It leads advocacy efforts to foster an inclusive environment for all identity groups, promoting diversity policies, organising seminars and events, Pride month, and collaborating with relevant stakeholders to support an equitable and welcoming research community. [See page 56.](#)

Organisation



Research Organisation

RT1 Probabilistic modelling and uncertainty

RT2 Structure and Logic

RT3 Scalable Approximation of Models and Algorithms

RT4 Integrative learning

RT5 Transfer learning

RT6 Causal learning

RT7 Natural language

RT8 Explainability and fairness

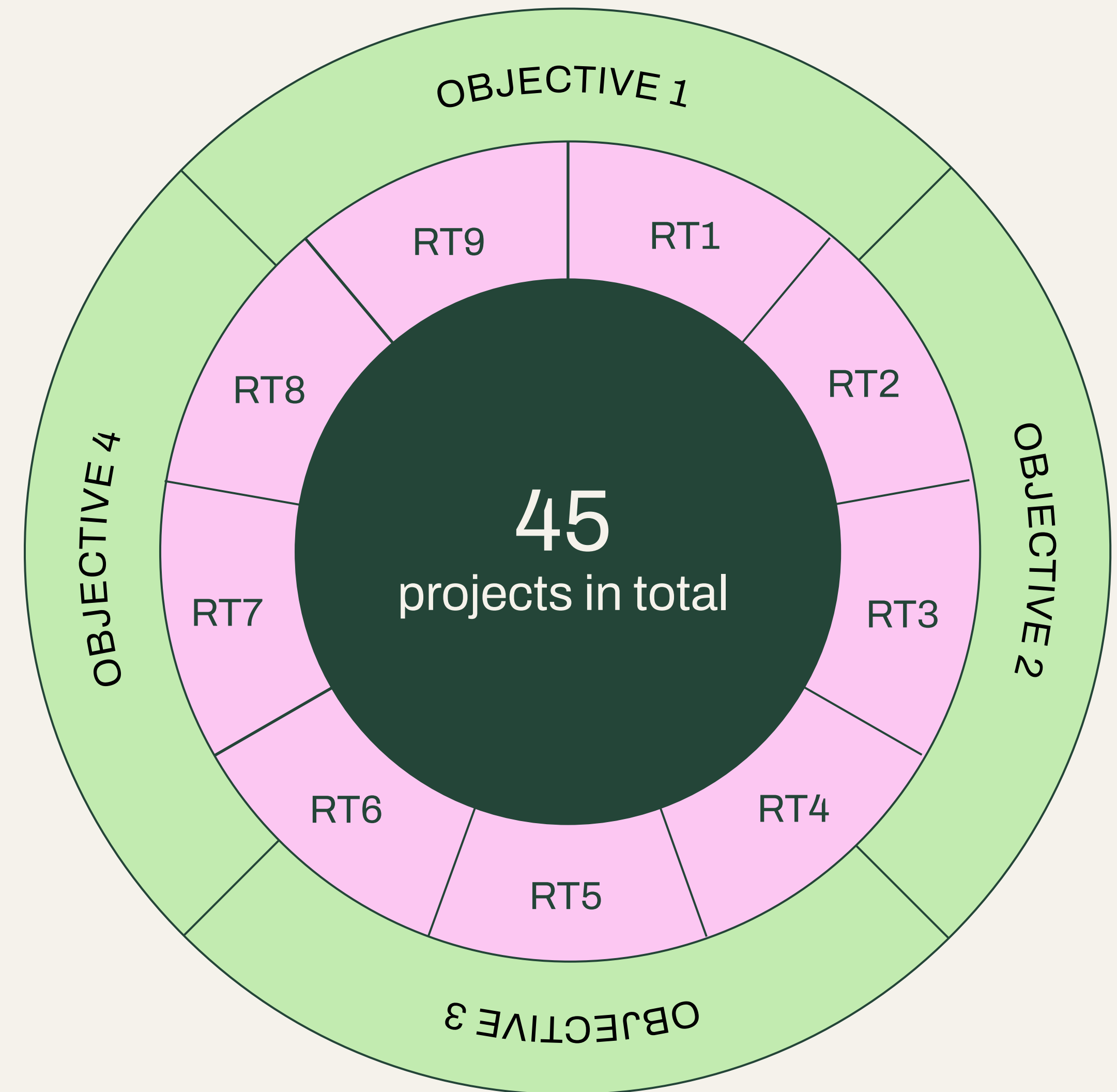
RT9 Ethics

Objective 1 Accurate knowledge-driven machine learning

Objective 2 Sustainable and green machine learning

Objective 3 Fair, explainable and trustworthy machine learning

Objective 4 Machine learning with quantified uncertainty



Objective 1: Accurate knowledge-driven machine learning

The usual way information is introduced into machine learning (ML) is implicit and by means of massive training data. Instead, Integreat will develop new methods to inject available knowledge directly into analyses, so to increase precision and robustness of solutions while shortcutting learning efforts. ML driven by domain-knowledge will leverage knowledge about the structures, processes and dynamics of the modelled system and about the data generating mechanisms. We will formulate new frameworks that combine exact domain knowledge (e.g., in the form of must-link and cannot relationships, logical formalisations and formal ontologies), more imprecise and subjective information (e.g., most-likely-links, prior beliefs and stochastic relations) and data, in a coherent probabilistic theory, leading to provably more accurate discoveries, predictions and decisions.

Furthermore, to increase accuracy of ML, we will study methods to enlarge training data on the basis of knowledge, which require understanding of the coverage space of data, tail behaviour, and optimal sample design. Smaller, noisier, or even biased data can become more useful when knowledge drives ML. Knowledge-driven ML will be designed to also work with such data.

Objective 2: Sustainable and green machine learning

The carbon impact of data-centric ML is growing. State-of-the-art ML algorithms can consume an immense amount of energy, and models are often trained on costly, large, manually-annotated data sets. To achieve a linear gain in performance, it is necessary to train an exponentially larger neural net.

The footprint of data-centric ML can to some extent be reduced by more efficient hardware and use of renewable energy, but not sufficiently to make it economically, technically, and environmentally sustainable in the future. Integreat will add a new dimension, developing knowledge-driven methods for saving energy. Our research will contribute to this aim, e.g., by reusing and integrating data, transfer learning, developing causal and logical models (as these persist) and developing more parsimonious models. We will document in detail the energy footprint of our algorithms and methods, to allow for more complete cost-benefit analyses. Because storage also consumes energy, Integreat will propose knowledge-driven ways to compress, project and reduce data, while managing the loss of information. Furthermore, we will develop active learning methods, to determine how training data could be enlarged systematically and sustainably. UN Sustainable Development Goals 12 and 13 are directly addressed by Integreat, and knowledge-driven ML will indirectly contribute to many of their other goals too.

Objective 3: Fair, explainable and trustworthy machine learning

Today, ML can perpetuate and even amplify discrimination, because models are trained on historical data, which are often biased and non-representative, and the data labels can reflect outdated policies. Integreat will develop knowledge-driven methods to de-bias data sets and models. We will further work towards theoretical guarantees of fairness, exploiting a priori domain knowledge. For example, we will characterise fairness of algorithm-based decisions, by representing how variables change in response to knowledge-driven counterfactual interventions and exploiting logic-based knowledge representations. Explainability of black-box algorithms, such as deep learning, is today a major challenge for ML. As ML systems begin making decisions previously entrusted to humans, it becomes critical to explain the reasons behind their conclusions and make them fully interpretable by humans. Current ML is unable to explain the rationale and reasons for its solutions, while Integreat's knowledge-based approaches will be explainable. Our new solutions will enable understanding of the inner workings of deep learning and other black- and grey-box approaches, by augmenting them with knowledge-based explainable white-box components. Integreat will examine the fairness and explainability of ML algorithms as a necessary component of building public trust in them. We will detect and solve the lack of robustness in traditional ML approaches, in order to achieve greater trustworthiness. ML algorithms are easily fooled into misclassifications by minor perturbations of the data and by out-of-distribution data.

The foundations of fairness, explainability and trustworthiness reside in ethical analysis. It is necessary to examine from a philosophical position how these concepts are defined and practiced, in which way researchers, including ourselves, face the dilemma between accuracy vs. transparency, and how our results are committed to respecting each individual. Ethics research in Integreat will develop solutions to these questions.

Objective 4: Machine learning with quantified uncertainty

Data are most often incomplete, noisy and inconsistent, knowledge and models are imperfect, and algorithms are approximate. Therefore, estimates, generalisations, predictions and decisions produced by ML are intrinsically uncertain. To provide the trust needed to inform decisions, uncertainty of results must be quantified. Current ML struggles to assess uncertainty in a precise and coherent way. Therefore, at Integreat we will represent and model multiple sources of uncertainty probabilistically, thus quantifying the reliability of results. Importantly, uncertainty quantification can also reveal disagreement between data and knowledge, as well as between different data sources. We will develop methods that automatically alert us of such discrepancies, a safety insurance for our knowledge-driven methods. At Integreat we can reduce the uncertainty of solutions considerably by combining knowledge and data.

Research highlights

Integreat's research is organised in nine interconnected research themes (RTs), each focusing on specific methods and challenges of knowledge-driven ML and each contributing to several of our objectives O1-O4. In this part of the annual report, we briefly describe the purpose and content of each RT. Of course, during the years of the centre, the RTs will adapt to the developments and disruptions of the field. Each RT embraces multiple projects, which in turn can relate to several RTs. Each RT is guided by a small team of Integreat PIs. Most of the projects involve a PhD fellow, or a postdoc.

The focus of the past 18 months has been on building Integreat, recruiting new researchers, postdocs, PhD fellows, and administrative staff. Research activity has been ramping up and first outputs are included in this report and on the Integreat website. The first results are emerging as preprints on arXiv, with some already accepted at ML conferences, workshops, and journals. While all publications collected in this annual report have benefited in various ways from Integreat, it is natural that several originate from work started before the centre's establishment. Our papers are

published in leading AI and ML conferences and journals, including one paper in each of ICML, ICLR, AISTATS, LICS, Nature Machine Intelligence, three papers in AAAI conference on AI and two papers in JMLR. In addition, we have published around ten papers in other ML conferences and workshops, 20 papers in ML and statistics journals and a few papers on domain specific journals.

A representative selection of our work includes:

- › A new framework for approximate Bayesian inference intended for a situation where only a limited number of noisy log-likelihood evaluations can be obtained, approximating the log-likelihood function using a Gaussian process.
- › A new approach for synthetic tabular data generation, which is superior to differential privacy. Instead of adding noise, it cancels conditional dependences involving features that are privacy protected, using knowledge-guided vine copula truncation.
- › We analyse the state of the art of prior knowledge elicitation, from properties of the modelling task and the nature of the priors to the form of interaction with the expert. Our extended literature review allows recognising understudied directions in prior elicitation research, leading to a proposal of several new avenues to improve prior elicitation methodology.
- › Similarly, we review software tools for Bayesian inference, in particular active development in the ML community. We provide a useful map to navigate the world of Bayesian computation. We anticipate a vigorous development of algorithms in probabilistic programming, likelihood-free inference and Bayesian neural networks.

- › We conducted extensive simulation studies to evaluate how different methods to estimate conditional Shapley values. We propose a new use of Shapley values for the attribution of data sources for prediction, in particular how different data-clusters of the training data affect a prediction.
- › We develop a new theoretical understanding on the generalisation error of Stochastic Gradient Descent with momentum and we propose a modified momentum-based update rule with early momentum.
- › We address label shift in distributed learning via entropy regularisation. We propose the Versatile Robust Label Shift method, which enhances the maximum likelihood estimation of the test-to-train label density ratio.
- › Grounded language models use external sources of information, such as knowledge graphs. By extending previous work on compositional generalisation in semantic parsing, we allow for a controlled evaluation of the degree to which these models learn and generalise from patterns in knowledge graphs.

In the upcoming year, work on the research themes will increase significantly, and we expect more results, some of groundbreaking character.

QR to the list of PIs



[Visit webpage](#)

QR to Research



[Visit webpage](#)

September 2023
→ March 2025

RT1: Probabilistic modelling and uncertainty



Visit webpage

Probabilistic ML equips ML algorithms with stochastic components and measures of uncertainty. Modern AI is rich of probabilistic elements, for example in most of generative AI, diffusion models, optimisation algorithms and conformal prediction. Large Language Models are intrinsically probabilistic, predicting the distribution of the next word in a text. The stochastic mechanisms allow to represent good approximations of real-world systems and to generate a realistic variety of synthetic outputs. Integreat will develop approaches that incorporate domain and a-priori knowledge in order

to increase the power and precision of ML models. We want to inject knowledge into ML by developing further the probabilistic, Bayesian, and post-Bayesian perspectives, through new forms of informative priors, penalties, complex latent structures, model ensembles, and network designs. Often, stationarity and equilibrium of systems cannot be assumed; new and different environments and data become available; knowledge changes because of new understanding or varying opinions, and we will detect such changes and exploit them to make better out-of-distribution predictions. We will develop theoretical guarantees of properties of new methods, in addition to empirical validation on benchmark data, cross-validation, and assessing predictability on new data. Too often benchmarks represent surpassed situations and may be a result of biased or unrealistic labelling. We will critically assess benchmark quality and generalisability by exploiting knowledge about the application domain.

Quantifying uncertainty in ML means to go beyond a pure point prediction, by estimating the probability of such a prediction. One wishes to determine the set of values which are predicted to happen with high probability. For example, an ML algorithm might classify a media article according to its topic, say, economy, politics, debate, sport, culture, etc. A probabilistic prediction would mean to assign the probability of each topic, say, 60% politics, 20% economy, 10% culture, 10% debate. A second level of uncertainty quantification is to further describe the uncertainty of these probabilities, for example one could estimate politics with between 55% and 67%, etc (seen as a subset of the unit simplex). There are methods to quantify uncertainty, but they are unprecise, do not include all components of uncertainty, and are difficult to use in situations of complex dependence. Because uncertainty is expected to shrink when good quality data grow or additional knowledge becomes available, Bayesian uncertainty quantification allows to detect the incapacity to learn further, also allowing to save energy.

Integreat will evaluate the usefulness of knowledge in probabilistic ML to reach our four objectives: which knowledge is relevant, as well as when, where, and how it should be infused into the models. We will study robustness and stability to knowledge assumptions, investigate the effects of false knowledge and divergence between knowledge, models, and data.

PROJECTS

Bayesian adaptive networks via Kernels and Nonparametrics

We enhance model interpretability and improve scalability and efficiency by integrating Gaussian processes and deep learning.

Data-driven tensor factorisation models

Tensor factorisation methods are widely used, but analysing tensor data presents significant challenges for latent variable modelling due to their high dimensionality, heterogeneity, and complex structure. We explore the use of ML techniques to enhance tensor-based modelling and propose a unified perspective on this diverse field. To highlight the significance of recent advancements in this area, we contributed to a discussion on Bayesian Analysis (2025, Vol. 20, No. 1, pp. 213–344) recognizing the importance of the paper *Sparse Bayesian Factor Analysis When the Number of Factors Is Unknown* by Sylvia Frühwirth-Schnatter and colleagues.

Discrepancy measures and accelerated likelihood-free inference for simulator-based models

We are using generalised Bayesian inference approach for updating beliefs of model parameters. The score function relating model parameters and observed data is approximated with probabilistic surrogates, such as Gaussian processes. Our interest lies primarily in the uncertainty that the probabilistic surrogates cause in importance sample posterior weights.

Embedded sufficient statistics

The information bottleneck is important in the construction of powerful ML algorithms that predict an output given an input. We study minimal statistical sufficiency to summarise the input in the most compressed way while maintaining prediction power.

Hybrid models for personalised medicine

By combining probabilistic mechanistic models of a system under study, for example a system of differential equations or stochastic cellular automata, with deep neural networks, we strengthen predictions and their uncertainty quantification.

Knowledge-informed spatio-temporal models

We build new, knowledge-informed probabilistic models for spatio-temporal data with a focus on scalability to high resolutions in space and time. The approach utilises and expands new classes of models for spatial extremes to develop efficient models that accurately capture data features in both the bulk and tail of the distribution.

Learning differential models (and other logic rules) from data with uncertainty

We train a deep neural network to estimate the system of differential equations that a multivariate time series follows, in symbolic form, and use this equation with its uncertainty quantification as a biomarker for personalised cancer treatment.

Learning from human preferences in Large Language Models

Modern large language models must be aligned with human needs. We develop a new approach to reinforcement learning from human feedback useful to align large language models with desired principles. The method uses advanced preference learning and the Mallows model, to describe the uncertainty of the consensus of the human (or algorithmic) evaluators. This will allow us to better aggregate human judgments and make use of uncertainties and disagreements between annotators.

Multi-modal data integration for personalised treatment recommendations

A theory of redundancy in multimodal data allows us to coherently impute only the needed missing components of data. For personalised treatment recommendations we develop logic-based decision rules and decision systems, where biological and clinical knowledge guides the search space exploration.

Probabilistic preference learning with the Mallows model

We incorporate knowledge about the items to guide recommendation based on users' expressions of preference. Recommendations are based on a new variational Bayesian approximation which allows scaling.

Sparse models for machine learning from a Bayesian viewpoint

We explore new methods to construct priors in Bayesian neural networks. We look at sparse models, modified horseshoe priors and combinations of neural networks and random forest methods.

Statistical models and logic: handling inconsistencies

We rethink models to handle inconsistency in data. Different treatment may be needed depending on whether inconsistencies are thought to arise due to aleatoric or epistemic uncertainty.

Statistically Optimal and Computationally Efficient Methods in Change-point Analysis

Computationally efficient methods are imperative for change and anomaly detection in data sequences. It is also critical to have control on the statistical properties of the detection procedures. We develop novel efficient methods in offline and online settings, for which we provide strong mathematical performance guarantees, which in several cases are shown to be optimal in a minimax rate sense.

Structure-generating models for transition metal complexes

The goal is prediction of the properties of transition metal complexes with uncertainty quantification, with possible extensions of current molecular structures.

Synthetic data generation balancing privacy and utility, using vine copulas

We develop TVineSynth, a vine copula based synthetic tabular data generator, which is designed to balance the protection of privacy against the utility. Unlike most existing work, we do not add noise globally, but instead perform targeted knowledge-driven privacy preserving model approximation, so that the utility of the resulting

synthetic data for downstream prediction tasks does not suffer.

Transforming real-world evidence with large language models: predictive analytics for improved healthcare outcomes

Can Large language models be fine-tuned on health registry data, so that they can predict aspects of the future health of individuals based on their longitudinal life-history? How should the fine-tuning be done, and how should the life history be summarised from tabular data in a prompt?

Uncertainty quantification in the presence of logical constraints

Conformal prediction (CP) is a common approach to quantify uncertainty of ML models. Graph Neural Networks (GNNs) is a family of ML architectures, that allows to naturally incorporate logical constraints. We design methods to improve the efficiency of CP for GNNs by taking such constraints into account.

What's in the Black-Box? Improving the Shapley Value Explanation Methodology

We have developed new and better ways of estimating Shapley values for XAI for dependent and mixed features based on Variational Autoencoders, and conditional Shapley values. We have compared a wide range of both novel and well-known estimation methods for Shapley values, studied the uncertainty of individual explanations, and improved sampling strategies.

RT1 TEAM LEADERS

- **Professor Fred Godtliebsen**,
UiT, Department of Mathematics and Statistics
- **Professor Thordis Thorarindsdottir**,
UiO, Department of Mathematics
- **Associate Professor Ingrid Hobæk Haff**,
UiO, Department of Mathematics
- **Associate Professor Ida Scheel**,
UiO, Department of Mathematics
- **Associate Professor Valeria Vitelli**, UiO, OCBE

RT1 TEAM

- **Associate Professor Ali Ramezani-Kebrya**,
UiO, Department of Informatics
- **Associate Professor Aliaksandr Hubin**,
UiO Department of Mathematics
- **Senior Researcher Alvaro Köhn-Luque**, OCBE
- **Research Director Anders Løland**, NR
- **Associate Professor Andrei Kutuzov**,
UiO, Department of Informatics
- **Associate Professor Anne-Marie George**,
UiO, Department of Informatics
- **Professor Arnaldo Frigessi**,
UiO, OCBE and Department of Mathematics

- **PhD Fellow August Arnstad**,
UiO, Department of Mathematics
- **Researcher Basil Ell**,
UiO, Department of Informatics
- **Associate Professor Benjamin Ricaud**,
UiT, Department of Physics and Technology
- **Researcher Celestino Creatore**, UiO, OCBE
- **PhD Fellow Christoffer Lingjærde**, UiO, OCBE
- **Professor Claudia Czado**,
Technical University of Munich
- **Associate Professor Egor Kostylev**,
UiO, Department of Informatics
- **PhD Fellow Elisabeth Maria Griesbauer**,
UiO, OCBE
- **Professor Erik Velldal**, UiO,
Department of Informatics
- **Professor Geir Olve Storvik**,
UiO, Department of Mathematics
- **Researcher Henri Pesonen**,
Oslo University Hospital, OCBE
- **PhD Fellow Håkon Tasken**, OCBE
- **Professor Ingrid K. Glad**,
UiO, Department of Mathematics

- **Associate Professor Johan Pensar**,
UiO, Department of Mathematics
- **Professor Jukka Corander**, UiO, OCBE
- **Research Director Kjersti Aas**, NR
- **PhD Fellow Lars Henry Berge Olsen**,
UiO, Department of Mathematics
- **Postdoctoral Fellow Leonard Schmiester**,
UiO, OCBE
- **Professor Lilja Øvrelid**,
UiO, Department of Informatics
- **PhD Fellow Lorenzo Zuccato**,
UiO, Department of Mathematics
- **PhD Fellow Luca Cibinel**,
UiO, Department of Mathematics
- **Professor Martin Giese**,
UiO, Department of Informatics
- **Senior Researcher Martin Jullum**, NR
- **Senior Researcher Martin Tveten**, NR
- **PhD Fellow Michele Chini**,
UiO, Department of Mathematics
- **Postdoctoral Fellow Nikolay Arefev**,
UiO, Department of Informatics

- **PhD Fellow Per August Jarval Moen**,
UiO, Department of Mathematics
- **Associate Professor Riccardo de Bin**,
UiO, Department of Mathematics
- **Professor Robert Jenssen**,
UiT, Department of Physics and Technology
- **PhD Fellow Sigurd Holmsen**,
UiO, Department of Informatics
- **Visiting Professor Sylvia Richardson**,
University of Cambridge, MRC Biostatistics Unit, OCBE
- **Professor Tero Aittokallio**, OCBE
- **PhD Fellow Vilde Hansteen Ung**, UiO, OCBE
- **Postdoctoral Fellow Vladislav Mikhailov**,
UiO, Department of Informatics
- **Researcher William Denault**,
Oslo University Hospital, OCBE
- **Postdoctoral Fellow Youness Azimzade**,
UiO, OCBE
- **Professor Øystein Sørensen**,
UiO, Department of Psychology

RT2: Structure and logic



Visit webpage

Most real-life data are structured. For example, they are often relational and stored in multiple interlinked database tables, hierarchical as in XML, or in the form of knowledge graphs in formats such as RDF and Property Graphs. Logical languages can be used to systematise and formalise complex domain knowledge, for example, in the form of ontologies, and logical inference can be used to enrich explicit structured data with implicit information that logically follows from the explicit data and formalised domain knowledge. Most common ML methods today can neither ingest nor produce structured data directly, and they are unable to take into account existing domain knowledge formalised in a logical language.

Currently, ML usually deals with structured data by employing ad-hoc embedding techniques and essentially re-learning the structure from training examples. Recent advances show that ML can learn the impact of logical knowledge indirectly or directly, for example, via embedding methods that incorporate the constraints expressed by the logical knowledge in the loss function. However, indirect approaches lead to difficulties in learning good quality models, unsustainable training (models need to be retrained when the logical knowledge changes), and poor interpretability. Among the most promising approaches that allow ML to cope with structure and logical knowledge are ontology embeddings, graph neural networks (GNNs) of various forms, and

representation learning in general. Combining these approaches with other existing methods (Inductive Logic Programming, Markov Logic Networks, ontology-based embeddings, etc.), Integreat develops new methods to intertwine ML with logical inference on deterministic and probabilistic knowledge in logical form.

Current work in Integreat has four focus areas:

- 1. Develop and study Graph Learning architectures (Egor Kostylev)
- 2. Derive logical knowledge from an ML model (Ana Ozaki)
- 3. Generalise sequence models to structure models (Basil Ell)
- 4. Develop similarity measures for graph structured data (Elisabeth Wetzner)

Graph Learning Architectures

One current line of Integreat’s work for integrating structured data into ML aims to consolidate the theoretical foundations of Graph Learning, in particular GNNs, and to investigate the practical applicability of such models. For example, we study the properties of GNN variants, which include establishing their expressive and approximating power, developing ways to incorporate various types of knowledge, and analysing their applicability to practical problems, such as knowledge graph completion, query answering over incomplete graphs, and explainability of neural models. There are

many GNN architectures for which seemingly small variations lead to dramatic changes in their properties. One of the key steps here is to design novel translation algorithms from GNN architectures to logical axioms and back. This work will allow us to bridge GNNs and classic logic-based Knowledge Representation formalisms in a principled way and give us an opportunity to leverage strong sides of both statistical and symbolic AI paradigms and to mitigate their weaknesses.

Deriving Logical Knowledge from ML Models

Large ML models can be overly complex to study without relying on ways of abstracting them. To better understand the information such models have learned, we will investigate active learning approaches for querying ML models, that is, approaches that probe ML models with a series of inputs and combine information from the outputs in a desired structured format. This includes the development of strategies for probing ML models, notably language models, to extract rules and decision trees that can help to detect trends and biases, in particular, to detect the presence of potentially harmful social biases. The improvement of question answering capabilities of language models in a variety of domains also facilitates the application of active learning approaches for knowledge acquisition. Building on these improved capabilities, we are interested in extracting ontologies representing domain knowledge present in trained language models.

Generalising Sequence Models to Structure Models

In another line of work, we focus on generalising sequence models, i.e., models that deal with sequences of tokens, such as language models, to structure models, i.e., models that can natively deal with graph-structured data. Where language models enable to estimate the probability of a sequence and to continue a started sequence, structure models will estimate the probability of a graph and continue/extend a graph. Structure models will be built on graph theory, representation learning, statistics, and logic. We envision that structure models can carry out both inductive and deductive reasoning, might more effectively include formally represented background knowledge, might be easier to update and more explainable than purely neural alternatives. We apply these models in the domain of chemistry and beyond.

Similarity Measures for Graph Structured Data

Another focus is learning similarity measures between labelled graphs of different sizes by incorporating additional knowledge about the relationship of features or labels that reside on their edges and nodes. One approach is to use representation learning techniques to learn representations of graph structures and to use non-Euclidean geometries to embed relational taxonomies of labels in which continuous distance measures between labels can be used to sample augmentations that integrate graph label information.

Similarly, useful knowledge embeddings may be found for edge feature relationships that can be incorporated in learning meaningful similarity measures for graph structured data.

PROJECTS

Compositional reasoning over graphs with Language models

This project combines two orthogonal research directions: compositionality in neural networks and the grounding of language models in external information sources, such as knowledge graphs.

Finding a middle ground based on lexicographic preferences

Based on a logic perspective of consistency and entailment, we define the notion of a middle ground to be a minimal set of coherent statements that are consistent with multiple stakeholders’ preferences. Under the assumption of lexicographic preference orders, we develop properties and algorithmic results w.r.t. the existence, uniqueness and construction of such a middle ground.

Structure-generating models for transition metal complexes

Our aim is to develop methods to use knowledge graphs (RDF). Currently, the information is added to the neural network in a very simplistic way, and our work will address this issue more thoughtfully.

Uncertainty quantification in the presence of logical constraints

Conformal prediction (CP) is a common approach to quantify the uncertainty of ML models. Graph Neural Networks (GNNs) is a family of ML architectures, that allows to naturally incorporate logical constraints. We design methods to improve the efficiency of CP for GNNs by taking such constraints into account.

Statistical models and logic: handling inconsistencies

Data can be viewed as a set of statements over a logical language. This allows us to define and identify inconsistencies in our data in new formal ways. We aim to transfer this notion of inconsistency to statistical models.

RT2 TEAM LEADERS

- **Researcher Basil Ell,**
UiO, Department of Informatics
- **Professor Martin Giese,**
UiO, Department of Informatics
- **Associate Professor Egor Kostylev,**
UiO, Department of Informatics
- **Associate Professor Ana Ozaki,**
UiO, Department of Informatics
- **Associate Professor Elisabeth Wetzer,**
UiT, Department of Physics and Technology

RT2 TEAM

- **Associate Professor Anne-Marie George,**
UiO, Department of Informatics
- **Professor Arild Waaler,**
UiO, Department of Informatics
- **Professor Arnoldo Frigessi,**
UiO, Department of Mathematics/OCBE
- **Associate Professor Ida Scheel,**
UiO, Department of Mathematics
- **Associate Professor Johan Pensar,**
UiO, Department of Mathematics
- **PhD Fellow Lorenzo Zuccato,**
UiO, Department of Mathematics
- **PhD Fellow Luca Cibinel,**
UiO, Department of Mathematics
- **Associate Professor Riccardo de Bin,**
UiO, Department of Mathematics
- **Associate Professor Valeria Vitelli,** UiO, OCBE

RT3: Scalable Approximation of Models and Algorithms



[Visit webpage](#)

In Bayesian inference, Markov Chain Monte Carlo (MCMC) algorithms are considered the gold standard but are far too slow for big data and complex models. Recent developments (from superefficient subsampling and quasi-stationary Monte Carlo (MC), to non-reversible processes) are computationally unfeasible in high dimensions. A different approach is to approximate the original model with one which can be efficiently sampled from.

Integreat will develop approximation theories that preserve structure and knowledge, by moving away from simplistic approximations, like mean-field models. In the spirit of normalising flows, we will propose knowledge-based, explainable Variational Bayes approximations, using causal graphs and vine copulae, which are flexible decompositions of any high dimensional model, capturing well also complex multivariate tail dependence. Their particular structure makes them computationally tractable. Similarly, we will investigate how to make active use of knowledge in dimension reduction, matrix factorisation and vector space embeddings, for example by informative penalisations. One further direction is extending sequential MC methods to high dimensions, combining efficient sampling with online updating of static parameters.

When first-principles models of the system exist, Approximate Bayesian Computation (ABC) approaches are very useful: being likelihood-free, they can be scalable, and because of relying on a mechanistic simulator,

they can preserve knowledge. Integreat will deepen simulator-based machine learning (ML) by efficiently combining first-principles models with deep learning, to form more theoretically plausible generative models. Integreat will host ELFI (Engine for Likelihood-Free Inference, [elfi.ai](#)) which is a leading open software inference platform for simulator-based models, and which efficiently uses Bayesian optimisation through emulator models. Inspired by our recent work that uses GP emulators

with bounded errors to replace the proposal function in MCMC, we will combine GP emulators and neural network architectures. This can be done by algorithms that efficiently use emulator-generated training data (cheap to forward simulate) in deep learning (flexible and accurate with sufficient training). Incorporating knowledge in this way, we will significantly improve the exploration-exploitation decisions of ABC. We will also develop knowledge-steered data sampling procedures to guide optimisation for better performance.

Huge data storages emerge as threats to our climate. Integreat will suggest new ways to approximate and compact data developing further statistical sufficiency and data distillation to incorporate domain knowledge. Learning what to forget, like humans do, also reduces storage needs. We will develop methods which use knowledge to decide which part of the information in the data that needs to be synthetised in appropriate summaries for future tasks or transfer learning.

Integreat will explore cloud computing ecosystems, as distributed computing implies agility and adaptability, and saves energy. Federated training and distributed inference will be knowledge-driven, and we will propose algorithms that can automatically telescope and scale up or down, depending on accuracy needs.

PROJECTS

Bayesian adaptive networks via Kernels and Nonparametrics

The project aims at enhancing model interpretability and improving scalability and efficiency through approximations of deep neural networks Gaussian processes, or the integration of the two.

Discrepancy measures and accelerated likelihood-free inference for simulator-based models

We efficiently train the probabilistic surrogates in our simulation-based GBI approach while simultaneously drawing a weighted sample from the approximate posterior. Surrogate training is carried out with a minimal number of samples using a novel active learning approach that chooses query points by minimising the total variability in the sample weights associated with specific parameter values. We developed a framework for approximate Bayesian inference intended for a situation where only a limited number of noisy log-likelihood evaluations can be obtained due to constraints on the available computational budget, which is becoming increasingly common for expensive simulator-based models. We model the log-likelihood function using a Gaussian process see

Embedded sufficient statistics

To capture information relevance within an input for prediction of an output, we consider the notion of approximate prediction sufficiency. We efficiently measure the level of sufficiency by estimating the mutual information of the representation of input and the output variable. The approximate sufficient model is then used for prediction.

Energy-efficient and scalable deep learning

We focus on large-scale and distributed deep learning in multi-GPU settings and develop efficient compression schemes to accelerate training significantly. Our layer-wise quantisation approximates noisy stochastic gradients with much smaller size in bits to improve scalability.

Exploration and control of the inner representation in generative AI models

We investigate the latent space of deep models to understand the structure of the latent space of large and smaller models. We study the approximation in latent space. Identifying similarities and differences between models of different sizes is a key to design more efficient models, smaller but with still high accuracy. It may allow a better understanding of the distillation principle.

Hybrid models for personalised medicine

Probabilistic mechanistic models of a system under study, sometimes referred to as digital twins, are first order approximations of the real-world. The level of

approximation, though very rough, might be sufficient for the purpose (for example short term predictions). To improve the precision of prediction and quantify uncertainty, the mechanistic model is complemented by a neural network. We study different ways to integrate these approaches.

Knowledge-informed spatio-temporal models

We build new, knowledge-informed probabilistic models for spatio-temporal data with a focus on scalability to high resolutions in space and time. The approach utilises and expands new classes of models for spatial extremes to develop efficient models that accurately capture data features in both the bulk and tail of the distribution.

Multi-modal data integration for personalised treatment recommendations

We develop logic-based decision rules and decision systems, where biological and clinical knowledge is explicitly used to build model approximations and in inference for guiding the exploration of the model parameters search space.

Sparse models for machine learning from a Bayesian viewpoint

In this project, we explore new methods for describing priors in Bayesian neural networks. We focus on sparse models, leading to lighter approximations of the functions represented by a network.

Statistical models and logic: handling inconsistencies

To apply a Bayesian approach for learning tasks with inconsistent data, we will consider efficient algorithmic solutions and scalable approximations.

RT3 TEAM LEADERS

- **Professor Jukka Corander**, UiO, OCBE

- **Professor Geir Olve Storvik**,
UiO, Department of Mathematics

RT3 TEAM

- **Associate Professor Ali Ramezani-Kebrya**,
UiO, Department of Informatics

- **Senior Researcher Alvaro Köhn-Luque**,
UiO, OCBE

- **Associate Professor Anne-Marie George**,
UiO, Department of Informatics

- **Professor Arnaldo Frigessi**,
UiO, Department of Mathematics/OCBE

- **PhD Fellow August Arnstad**,
UiO, Department of Mathematics

- **Associate Professor Benjamin Ricaud**,
UiT, Department of Physics and Technology

- **PhD Fellow Daniel Kaiser**,
UiT, Department of Physics and Technology

- **Associate Professor Egor Kostylev**,
UiO, Department of Informatics

- **Professor Fred Godtliebsen**,
UiT, Department of Mathematics and Statistics

- **Professor Geir Olve Storvik**,
UiO, Department of Mathematics

- **Researcher Henri Pesonen**,
Oslo University Hospital, OCBE

- **Associate Professor Ida Scheel**,
UiO, Department of Mathematics

- **Postdoctoral Fellow Leonard Schmiester**,
UiO, OCBE

- **PhD Fellow Lorenzo Zuccato**,
UiO, Department of Mathematics

- **Professor Martin Giese**,
UiO, Department of Informatics

- **Professor Robert Jenssen**,
UiT, Department of Physics and Technology

- **PhD Fellow Sigurd Holmsen**,
UiO, Department of Informatics

- **Professor Sylvia Richardson**,
University of Cambridge, MRC Biostatistics Unit

- **Associate Professor Valeria Vitelli**, UiO, OCBE

- **PhD Fellow Vilde Hansteen Ung**, UiO, OCBE

- **Postdoctoral Fellow Youness Azimzade**,
UiO, OCBE



*The fun part of discussing language models with philosophers and statisticians is that you constantly **get to question** your own assumptions.*

– PI Erik Velldal, Department of Informatics

RT4: Integrative learning

Integreat will integrate different forms of knowledge into joint analyses, as well as knowledge with a large variety of data types. Knowledge can be expressed by means of stochastic models, differential equations, semantic technologies, graphical representations, and logic-based ontologies and so on. The present gap between these approaches is considerable, but Integreat will study how they can be used in combination.

Data can be big, heterogeneous and with variable granularity and signal-to-noise levels, smaller and curated, in the form of natural language texts, images, and data streams. Using these together is often difficult. With integrative learning we can increase accuracy and explainability, quantify uncertainty, and disentangle the incremental contribution made by each knowledge and data component to the results, while protecting the results against excessive confidence only due to multiple use of the same information. We will use constrained optimisation and regularisation (where penalisations carry prescribed knowledge), information theory, ontology-based data integration, and Bayesian hierarchical models, which so far have not been studied in combination for the purpose of knowledge and data integration.

A different challenge of integrative learning emerges when two or more data or knowledge sources carry inconsistent information about the unknown parameters

and processes. Integreat will exploit domain knowledge to trust data sources and knowledge unevenly, for example by automatically down-weighting data that show anomalies or that contradict knowledge more than measurement error would allow. A further version of data integration is active learning, where the algorithm in charge of the data analysis and decisions determines when and how to enrich its training data. Integreat will develop knowledge-based approaches to determine how to enrich existing data, e.g., by adding new layers of data and new types of measurements of the system under study, for optimal predictive accuracy.

Compared to the current approach of knowledge-agnostic data generation, knowledge-steered data enrichment and optimal experimental design will increase accuracy, improve decisions and allow energy saving. This is important in particular when learning multi-way interactions, where the number of combinations grows exponentially with the number of interacting objects, as e.g., in reinforcement learning. In multi-way interaction models, we will for example, investigate multi-source input data assuming a higher-order tensor structure.

PROJECTS

Data-driven tensor factorisation models

Tensors are important in multimodal integrative modelling. Tensor factorisation methods are widely used, but

analysing tensor data presents significant challenges for latent variable modelling due to their high dimensionality, heterogeneity, and complex structure. We explore the use of ML techniques to enhance tensor-based modelling and propose a unified perspective on this diverse field. To highlight the significance of recent advancements in this area, we contributed to a discussion on Bayesian Analysis (2025, Vol. 20, No. 1, pp. 213–344) recognising the importance of the paper Sparse Bayesian Factor Analysis When the Number of Factors Is Unknown by Sylvia Frühwirth-Schnatter and colleagues.

Exploration and control of the inner representation in generative AI models

We are designing new interfaces to enable a better flow of information between a vision deep model and a language model, using the latent representation of the vision model. While progressing at a fast pace, language models still struggle to analyse vision token efficiently. Also, vision tokens may take a large part of the LLM context and we aim at reducing that. We are also working on creating a benchmark data set that investigates the use of the context in a chain-of-thought.

Hybrid models for personalised medicine

Probabilistic mechanistic models of a system under study, sometimes referred to as digital twins, are first order approximations of the real-world. The level of



[Visit webpage](#)

approximation, though very rough, might be sufficient for the purpose (for example short term predictions). To improve the precision of prediction and quantify uncertainty, the mechanistic model is complemented by a neural network. We study different ways to integrate these approaches.

Multi-modal data integration for personalised treatment recommendations

When integrating multimodal data, a theory of redundancy across modalities needs to be developed, so that one can exploit the redundancy to coherently impute only the needed missing components of data. We contribute to the definition of redundancy across modalities, so that critical aspects connected to missing data (e.g. block missing) can be overcome.

Structure-generating models for transition metal complexes

The final goal is to integrate knowledge (in the form of Resource Description Framework (RDF) graphs) into a prediction model. Currently we are adding information from an RDF graph into a graph neural network originally developed for a classical graph representation.

Transforming real-world evidence with large language models: predictive analytics for improved healthcare outcomes

We fine-tune Large Language Models using health

registry data, which include multiple layers of information, including longitudinal time series at very different scales, text, tabular data and possibly images. We study the integration of these components in training and in inference.

RT4 TEAM LEADERS

- **Associate Professor Riccardo de Bin**,
UiO, Department of Mathematics
- **Professor Tero Aittokallio**,
Oslo University Hospital,
Institute for Cancer Research

RT4 TEAM

- **Associate Professor Ali Ramezani-Kebrya**,
UiO, Department of Informatics
- **Associate Professor Aliaksandr Hubin**,
UiO, Department of Mathematics
- **Senior Researcher Alvaro Köhn-Luque**,
UiO, OCBE
- **Professor Arnaldo Frigessi**,
UiO, Department of Mathematics/OCBE
- **Researcher Basil Ell**,
UiO, Department of Informatics
- **Associate Professor Benjamin Ricaud**,
UiT, Department of Physics and Technology

- **PhD Fellow Christoffer Lingjærde**, UiO, OCBE
- **PhD Fellow Daniel Kaiser**,
UiT, Department of Physics and Technology
- **Associate Professor Johan Pensar**,
UiO, Department of Mathematics
- **Postdoctoral Fellow Leonard Schmiester**,
UiO, OCBE
- **PhD Fellow Luca Cibinel**,
UiO, Department of Mathematics
- **PhD Fellow Michele Chini**,
UiO, Department of Mathematics
- **Professor Sylvia Richardson**,
University of Cambridge, MRC Biostatistics Unit
- **Professor Tero Aittokallio**,
Oslo University Hospital,
Institute for Cancer Research
- **Associate Professor Valeria Vitelli**, UiO, OCBE
- **Researcher William Denault**,
Oslo University Hospital, OCBE
- **Postdoctoral Fellow Youness Azimzade**,
UiO, OCBE

RT5: Transfer learning



[Visit webpage](#)

Transfer learning has led to ground-breaking progress in natural language processing, computer vision and many application domains, with deep models pre-trained on massive amounts of unlabelled data. The representation of data learned by self-supervised AI models provides a powerful basis for downstream machine learning tasks. Moreover, pre-training has become the default practice in deep learning to substantially save compute costs and data. Pre-trained encoders (also called foundation models) compress raw data into compact latent embeddings that preserve rich semantic information. Understanding distribution and structure of these embeddings is an important problem. Meanwhile, sharing these embeddings without adequate protection exposes them to unauthorised use in a wide range of malicious activities.

Integreat will explore and analyse the inner representation of deep models, including language models. Data-driven approaches leading to these meaningful latent representations can be enhanced by combining knowledge-driven approaches on top of them. While the self-supervised learning of current language models, with next token prediction, targets superficial and local properties of text (e.g., word-in-context prediction), Integreat will study linguistically informed self-supervision, e.g., by estimation of latent variables based on prior knowledge, like lexical relationships between words. The representation of multimodal data

is also of key importance with the question of how to interface language models with images, audio or video. Self-supervised learning is not enough to create efficient connections between modalities and Integreat aims at going beyond this paradigm.

Integreat will enable transfer learning across a broader set of domains and modalities by explicitly utilising prior knowledge of the domains and of the data generation processes, and by developing statistical and information-theoretic principles to ensure knowledge coherence. We will study new aspects of reproducing kernel Hilbert spaces, to match distributions and representations across tasks, and to gain in efficiency, interpretability and sustainability. Further, we will facilitate joint extraction of knowledge (soft and hard) across the domains by investigating domain alignment, which is crucial to combining multi-modal data, and provide knowledge-conditioned alignment solutions. Throughout, it will be important to quantify uncertainty of transferability to guide transfer learning across sources.

PROJECTS

Data-driven tensor factorisation models

Tensor factorisation methods produce embeddings in latent space, which can be used widely, but analysing tensor data presents significant challenges for latent variable modelling due to their high dimensionality, heterogeneity, and complex structure. We explore the use of ML techniques to enhance tensor-based

modelling and propose a unified perspective on this diverse field.

Distributional shift

We minimise overall true risk in distributed settings under label shifts from train to test time. We improve test-to-train density ratio estimation using Shannon entropy-based regularisation. Our method encourages smoother and more reliable predictions that account for inherent uncertainty and achieves up to 20% improvement in average test error in imbalanced settings. The results will be published at ICLR 2025.

Exploration and control of the inner representation in generative AI models

The latent representation of deep models is a central element of transfer learning. A better understanding of latent space can unlock better transfer of information from (few-shot, zero-shot learning) and between deep models (multimodality). We have focused on the transfer of information between a vision model and an LLM and intend to investigate other modalities over the next years.

Finding a middle ground based on lexicographic preferences

With an application focus on preferences in moral dilemmas, particularly the trolley problem as faced in the Moral Machine Experiment, this project offers an alternative approach to AI alignment. Instead of aggregating moral

preferences and forcing compromise which might be unacceptable for moral values, we aim at identifying a middle ground, i.e., a (sub)set of agreeable decisions.

On the Intersection between Information Theory and Deep Learning

We exploit the inherent geometry of neural networks with piecewise linear activation functions to provide natural bins for mutual information estimation. The resulting estimates can provide valuable insights into the inner workings of deep learning algorithms, which can play an important role in understanding the large language models and provide new perspectives on uncertainty modelling and explainability.

Theoretical understanding of deep neural networks

We showed that it is crucial to establish an appropriate balance between the optimisation error with regards to empirical risk and the generalisation error (gap between true risk and empirical one) when accelerating stochastic gradient descent. The results are published at JMLR.

RT5 TEAM LEADERS

- **Professor Michael Kampffmeyer,**
UiT, Department of Physics and Technology
- **Associate Professor Ali Ramezani-Kebrya,**
UiO, Department of Informatics
- **Associate Professor Benjamin Ricaud,**
UiT, Department of Physics and Technology

RT5 TEAM

- **Associate Professor Aliaksandr Hubin,**
UiO, Department of Mathematics
- **Associate Professor Ana Ozaki,**
UiO, Department of Informatics
- **Associate Professor Anne-Marie George,**
UiO, Department of Informatics
- **Professor Arnaldo Frigessi,**
UiO, Department of Mathematics/OCBE
- **Postdoctoral Research Fellow Changkyu Choi,**
UiT, Department of Physics and Technology
- **Professor Christos Dimitrakakis,**
Université de Neuchâtel
- **PhD Fellow Daniel Kaiser,**
UiT, Department of Physics and Technology
- **PhD Fellow Elisabeth Maria Griesbauer,**
UiO, OCBE
- **Associate Professor Elisabeth Wetzer,**
UiT, Department of Physics and Technology
- **Researcher Henri Pesonen,**
Oslo University Hospital, OCBE
- **Associate Professor Ingrid Hobæk Haff,**
UiO, Department of Mathematics
- **Professor Jukka Corander,** UiO, OCBE

- **Professor Robert Jenssen,**
UiT, Department of Physics and Technology
- **Professor Sylvia Richardson,**
University of Cambridge, MRC Biostatistics Unit
- **Professor Tero Aittokallio,**
Oslo University Hospital, Institute for Cancer Research
- **Associate Professor Valeria Vitelli,** UiO, OCBE
- **PhD Fellow Vilde Hansteen Ung,** UiO, OCBE
- **Researcher William Denault,**
Oslo University Hospital, OCBE
- **PhD Fellow Zhiyuan Wu,**
UiO, Department of Informatics

RT 6: Causal learning



[Visit webpage](#)

When humans reason, they typically think in terms of cause and effect, an ability that current ML models generally lack. This may prove to be a fundamental limitation; in fact, the lack of robustness and generalisation in current ML models has recently been linked to their incapacity to identify and exploit causal structures. One of the key characteristics of causal relationships is that they are stable (or even invariant) across domains. Hence, when an ML model is guided to capture causal relationships, it is more likely to generalise well to different contexts and scenarios outside its specific training domain. Moreover, causality-aware models are also thought to be more parsimonious than pure correlation-based models, enjoying an improved interpretability and sample efficiency.

Assuming that the system under study is modular and built up of independent causal components, a change in domain can be represented as a change in some of the causal components, while the remaining part of the system stays unchanged and can thus be transferred across the considered domains. Exploiting this invariance principle, Integreat will develop new methods for integrating causal knowledge into ML models as part of the training procedure. We are going to develop causality-aware ML methods that: (i) account for the degree of invariance given a diverse set of training data from multiple domains, (ii) impose existing causal knowledge in the form of a (partial) causal structure.

For situations where causal knowledge is lacking, Integreat will develop new methods for extracting such

knowledge from data, building upon the framework of structural causal models. Particularly, we are going to develop new methods for causal discovery through structure learning, a technique that has recently shown great potential when approached from a Bayesian perspective. To help guide the structure learning algorithm, and achieve identifiability also beyond the Markov equivalence class, we are going to consider intervention experiments and exploit the availability of a knowledge-based probabilistic models of the system to propose adaptive intervention designs.

PROJECTS

Causal multi-task machine Learning for Enhanced Versatility, Equity and Robustness

The accuracy of ML models might degrade considerably when applied outside their specific training domains. By leveraging ideas and principles from the field of causality, we develop new training procedures for improving the robustness of ML models to various domain shifts.

Causal discovery using directed graphical models

Understanding the behaviour of a system under the influence of external interventions is the ultimate goal of many scientific studies, and it requires knowledge about the underlying causal structure. We develop new scalable and uncertainty-aware methods for detecting causal relationships from data by combining structure learning and causal inference.

Structure-generating models for transition metal complexes

The goal is prediction of the properties of transition metal complexes with uncertainty quantification. Causal learning may be a possible approach. For example, we are fitting a DAG (ZiG-DAG) on the results of RDF pattern mining.

RT6 LEADERS

- **Associate Professor Johan Pensar**,
UiO, Department of Mathematics
- **Professor Arnaldo Frigessi**,
UiO, Department of Mathematics/OCBE

RT6 Team

- **Postdoctoral Fellow Camilla Lingjærde**,
UiO, Department of Mathematics
- **Professor Geir Kjetil Sandve**,
UiO, Department of Informatics
- **PhD Fellow Johan de Aguas**,
UiO, Department of Mathematics
- **PhD Fellow Luca Cibinel**,
UiO, Department of Mathematics
- **Associate Professor Riccardo de Bin**,
UiO, Department of Mathematics
- **Professor Sylvia Richardson**,
University of Cambridge, MRC Biostatistics Unit
- **PhD Fellow Vera Haugen Kvisgaard**,
UiO, Department of Mathematics

RT7 Natural language



[Visit webpage](#)

Recent breakthroughs in Natural Language Processing (NLP), powered by pre-trained large language models (LLMs), have transformed the landscape of AI. These models have demonstrated impressive capabilities across a wide range of applications by leveraging vast amounts of data and sophisticated neural architectures, showing great promise in enabling natural language as an interface between humans and machines. However, despite their successes, current LLMs exhibit critical limitations: they often lack robustness, produce biased or factually incorrect outputs, struggle with reasoning, and operate as "black boxes" with limited explainability. Furthermore, their development raises concerns about environmental sustainability, due to massive computational demands.

The current dominating LLM paradigm of trying to learn everything through next-token prediction alone, makes for models that are brittle, difficult to control, and extremely resource-demanding in terms of both data and compute. This motivates the need for research on incorporating knowledge into various stages of the LLM development cycle, from data preparation to pre-training, post-training, and inference, and, last but not least, evaluation. As with all machine-learning, LLM development is driven by a feedback loop of training and evaluation. However, given the generality of LLMs, the task for evaluating them is monumental; there are many

domains of knowledge and many possible modes of problem-solving to cover, not the least with respect to ethical dimensions, and the output space, in terms of both predictions and ground truth, is often open-ended.

In the settings above, relevant knowledge can come in various forms, like linguistic knowledge, world knowledge, domain-specific knowledge, etc., and can be represented in various ways, e.g. human annotations of text, knowledge graphs, databases, logical rules, and more. For example, developing more linguistically informed training objectives can enable self-supervised learning that is more data efficient than the current paradigm. Developing mechanisms that allow models to access and reason with external knowledge sources during prediction, can enhance their ability to explain decisions and adapt to new scenarios. There are also many open research questions pertaining to interpretability and explainability for LLMs.

Several activities spring out from the ongoing Integreat reading group on preference optimisation and alignment of LLMs. This topic has important research questions pertaining to both modelling and data creation that benefit from competence in across several areas – statistics, NLP, ethics, and more – making it a very good fit for Integreat. The regular meetings of the reading group have attracted a lot of interest, from both core members of Integreat and more loosely affiliated research

ers from associated research groups. The discussions in the reading group have resulted in concrete research plans that we will now take action to implement. One of these directions involves assessing LLMs' abilities to reproduce human preferences and justify their choices.

PROJECTS

Benchmarking ethical reasoning in Large Language Models

So-called benchmarking, i.e. the systematic and comparative evaluation of the capabilities, knowledge and shortcomings of models, is an essential part of developing LLMs. In this interdisciplinary effort, we seek to extend LLM benchmarking methodology to the domain of ethical reasoning and moral judgement, though from a descriptive rather than normative angle. This includes a project on assessing ethical and moral judgments by LLMs.

Causal multi-task machine Learning for Enhanced Versatility, Equity and Robustness

The accuracy of ML models might degrade considerably when applied outside their specific training domains. By leveraging ideas and principles from the field of causality, we develop new training procedures for improving the robustness of machine learning models to various domain shifts.

Compositional reasoning over graphs with LLMs

Central to human reasoning and language is the ability to form novel combinations from seen components, often referred to as compositional generalisation. This project investigates to what extent language models exhibit compositional generalisation, in particular when making use of external and structured knowledge. This project is an example of the synergies between Integreat and the Language Technology Group at IFI, UiO.

Knowledge and bias extraction from LLMs

We investigate the extraction of rules expressing knowledge and social bias from large language models. We employ techniques based on computational learning theory and study which theoretical guarantees can be given for the correctness of the rules extracted.

Learning differential models (and other logic rules) from data with uncertainty

To train a deep neural network to estimate the system of differential equations that a multivariate time series follows in symbolic form, we produce a unique new data base with all differential equations published in the world. We use LLMs to find these equations in scientific papers and to interpret and tag them.

Learning from human preferences in Large Language Models

Modern LLMs must be aligned to human needs by what is called human preference learning. This project aims to improve this stage by using a Mallows model from statistics, which is context-dependent. It will allow to better aggregate human judgments and make use of uncertainties and disagreements between annotators

Transforming real-world evidence with large language models: predictive analytics for improved healthcare outcomes

We fine-tune LLMs using health registry data, which include multiple layers of information, including longitudinal time series at very different scales, text, tabular data and possibly images. We study how to summarise these data into efficient prompts for training and inference.

RT7 TEAM LEADERS

- **Associate Professor Andrei Kutuzov**,
UiO, Department of Informatics
- **Professor Erik Velldal**,
UiO, Department of Informatics

RT7 TEAM

- **Senior Researcher Alvaro Köhn-Luque**,
UiO, OCBE
- **Associate Professor Ana Ozaki**,
UiO, Department of Informatics
- **Professor Anna Smajdor**,
UiO, Department of Philosophy,
Classics, History of Art and Ideas
- **Associate Professor Anne-Marie George**,
UiO, Department of Informatics
- **Professor Arnaldo Frigessi**,
UiO, Department of Mathematics/OCBE
- **Associate Professor Benjamin Ricaud**,
UiT, Department of Physics and Technology

- **PhD Fellow Emil Poiesz**,
UiO, Department of Informatics
- **PhD Fellow Hannah Clausen**,
UiO, Department of Informatics
- **PhD Fellow Håkon Tasken**, UiO, OCBE
- **Associate Professor Ida Scheel**,
UiO, Department of Mathematics
- **Professor Lilja Øvrelid**,
UiO, Department of Informatics
- **Professor Manuela Zucknick**, UiO, OCBE
- **Professor Martin Giese**,
UiO, Department of Informatics
- **PhD Fellow Michele Chini**,
UiO, Department of Mathematics
- **Postdoctoral Fellow Nikolay Arefev**,
UiO, Department of Informatics
- **Associate Professor Samia Touileb**,
UiB Department of Information Science
and Media Studies
- **PhD Fellow Sondre Wold**,
UiO, Department of Informatics
- **Associate Professor Valeria Vitelli**, UiO, OCBE
- **Postdoctoral Fellow Vladislav Mikhailov**,
UiO, Department of Informatics

RT8: Explainability and fairness



[Visit webpage](#)

We advance methods for explainability and fairness in complex AI systems. Explainability in AI helps us understand how AI makes decisions, ensuring it is transparent and trustworthy. Further, it is crucial to assess the fairness of such decisions where humans are involved and guarantee non-discriminative, equal treatment of all.

Deep learning is a powerful way of approximating functions. However, its solutions are often “black boxes”, which prevents their use in life- and safety-critical applications in health, transport, law and welfare. “White and grey” statistical models, such as generalised additive and vine copula models, are explainable, allow non-linear effects and complex dependence, but sacrifice complex interactions for interpretability.

Integreat will propose new methods to couple statistical models with deep learning, in order to exploit advantages from both. We will blend statistical sampling approaches with backpropagation-based feature explainability mechanisms, and study regularisations which imposes interpretability. Integreat will work on attributing the significance to each factor in explaining individual predictions taken by ML algorithms.

Methods we will develop originate from classical statistical testing, in-silico controlled knock-outs and game

theory, like Shapley values, which we have started to extend to dependent features. We believe that the currently most promising approach is using non-parametric vine copulae. A further important element of trustworthiness of ML models is robustness to model assumptions and inference, as well as to adversarial attacks. We will leverage knowledge about specific artefacts (e.g., adversarial backdoor attacks, Clever Hans effects), to develop robust systems that are self-explanatory. We will develop new ways to understand deep learning systems, utilising information theoretic concepts and knowledge-infused learning. The general limits of explainability in “black boxes” and the role of uncertainty quantification to provide explainability are further themes of research for us.

Fairness considerations are not to be neglected when making decisions for humans. If not taken into account, biases may persist or be amplified leading to undesirable or unfair decisions in our societies. We will propose approaches that detect and control the bias, starting from counterfactual fairness, that explicitly intervenes on protected features to quantify a potential bias. Furthermore, to determine which counterfactuals to test, we will detect factors that influence large numbers of variables in the causal graph representing the domain knowledge. We also investigate bias dynamics and

different policies, e.g. affirmative actions, and action recommendations, e.g., recourse, which may change the population over time. Here, we study their long-term effects on fairness.

Another research direction explores fairness guarantees for decisions made under uncertainty of agent’s preferences over decision outcomes.

PROJECTS

Knowledge and bias extraction from Large Language Models

We investigate the extraction of rules expressing knowledge and social bias from large language models. We employ techniques based on computational learning theory and study which theoretical guarantees can be given for the correctness of the rules extracted.

Synthetic data generation balancing privacy and utility, using vine copulas

We develop a vine copula based synthetic tabular data generator, which is designed to balance the protection of privacy against the utility. The way we control privacy is explainable: unlike most existing work, we do not add noise globally, but instead perform targeted privacy preserving model approximation, so that the utility of the resulting synthetic data for downstream prediction tasks does not suffer.

What’s in the Black-Box? Improving the Shapley Value Explanation Methodology

We have investigated the local, model-agnostic Shapley value framework for explaining black box regression type machine learning methods, in various realistic, complex situations where the standard estimation methods are not valid, like when features are statistically dependent. We have developed new methodology, and contribute significantly to the R and Python packages SHAPR and KernelSHAP.

RT8 TEAM LEADERS

- **Research Director Kjersti Aas**, NR
- **Associate Professor Anne-Marie George**, UiO, Department of Informatics
- **Associate Professor Kristoffer Wickstrøm**, UiT, Department of Physics and Technology

RT8 TEAM

- **Associate Professor Ana Ozaki**, UiO, Department of Informatics
- **Research Director Anders Løland**, NR
- **Professor Arnaldo Frigessi**, UiO, Department of Mathematics/OCBE
- **Researcher Celestino Creatore**, UiO, OCBE
- **Professor Christos Dimitrakakis**, Université de Neuchâtel
- **Professor Claudia Czado**, Technical University of Munich
- **PhD Fellow Elisabeth Maria Griesbauer**, UiO, OCBE
- **PhD Fellow Emil Poiesz**, UiO, Department of Informatics
- **Professor Erik Velldal**, UiO, Department of Informatics

- **Associate Professor Ingrid Hobæk Haff**, UiO, Department of Mathematics
- **Professor Ingrid K. Glad**, UiO, Department of Mathematics
- **Associate Professor Johan Pensar**, UiO, Department of Mathematics
- **PhD Fellow Lars Henry Berge Olsen**, UiO, Department of Mathematics
- **Professor Lilja Øvrelid**, UiO, Department of Informatics
- **Senior Researcher Martin Jullum**, NR

RT9: Ethics



[Visit webpage](#)

We analyse ethical aspects of knowledge-driven ML. We employ the methods of analytical philosophy, in conjunction with experimental philosophy and conceptual engineering, to examine three interrelated themes which pertain to all other research themes and define the ethical ground of our four objectives.

We think of all ethical questions as being multi-dimensional with multiple utilities that need to be defined clearly for research projects to be self-consistent. At the same time, all ethical questions must be contextualised and adaptation from one context to another is a difficult problem. Research areas such as transfer learning and AI alignment are key to tackle these challenges.

AI and reification

Integreat examines the ways in which AI relies on the capacity to reduce phenomena to what can be digitised, measured, and fed into algorithms. This process can be understood as a form of reification – the process by which a phenomenon becomes a ‘thing’. This may apply both to concepts, such as privacy, bias, consent, fairness, and to the inhabitants of the world: people are reified when classified in certain ways. Some ethical requirements tend to enforce this kind of reification: in order to measure bias, we need to ascribe labels to data. Our research examines the ethical implications of translating human behaviours and characteristics and values into quantifiable data points. We will explore

how this reification feeds back into understandings of ethical concepts. We will also investigate whether reification is a necessary aspect of AI, and the possibility of alternative approaches.

AI and autonomy

With greater reliance on machine learning (ML) in situations of everyday life, individuals are increasingly likely to be influenced in their decisions and choices. The principle of respect for autonomy is a key ethical concept here. Autonomy entails that individuals are able to make informed choices and act upon them in accordance with their beliefs and values. It is this that makes coercion so obviously unacceptable. However, the nature of ML algorithms means that softer forms of manipulation are frequently at play, and in contexts where people lack awareness of how they are being influenced. Integreat’s research aims to identify the demarcation between ethically-acceptable nudging and unacceptable manipulation. We look at the question of how reward functions play into the nudging capacity of AI systems, and consider when users have a right to know these reward functions.

An additional concern for autonomy in the context of AI revolves around the so-called ‘black box’ problem. Current literature suggests a dichotomy between transparency and reliabilism. Those who advocate transparency believe this is essential for trust in AI systems;

and legislative efforts, e.g. the EU AI Act, have therefore had a focus on transparency. Some argue that there is a “right to an explanation”. Others claim that transparency is simply not achievable. Conversely, “reliabilists” assert that if systems are accurate enough, transparency may be less important. Designers of ML systems thus face a dilemma: favour accuracy over explicability? How much can XAI (explainable AI) methods help to make black-box ML models more transparent? Integreat will ascertain what role transparency can and should play in our understanding of trust.

Justice and technological imagination

Justice concerns the fair distribution of burdens and benefits. But interpretations of these concepts vary according to context and culture. How do we conceptualise a just future with AI? What goals motivate researchers, and how do they relate to the values and concerns of the public and other actors? Who do we expect to benefit, and by what means? Epistemic injustice is an important consideration here. Our societies tend to privilege some voices over others; this feeds into the direction and aims of research in ways that are often rendered invisible. The way we imagine our technological futures has a profound impact on which pathways we pursue, and which possibilities we can identify as being worthwhile. These are questions that may be overlooked in the rush towards developing effective

solutions for problems identified as high-priority. Our research includes an exploration of the values that lead researchers and scientists working on AI. We ask how they imagine the future of our society with AI, and how are these visions materialised (or not) in their practice and what impediments they experience.

PROJECTS

Benchmarking ethical reasoning in LLMs

LLMs can be used to make predictions about outcomes that imply making ethical and moral judgements. Without anthropomorphising the underlying processes, we seek to better understand the moral reasoning in LLMs by repurposing methodology from experimental philosophy, moral psychology, and social science.

Learning from human preferences in LLMs

Modern large language models are sometimes required to be aligned to human needs or ethical principles, by what is called human preference learning. This project aims to improve this stage by using a Mallows model from statistics, which is context-dependent. It will allow to better aggregate human judgments and make use of uncertainties and disagreements between annotators. This is to be expected because what is “more ethical” is an individual opinion.

Knowledge and bias extraction from LLMs

We investigate the extraction of rules expressing knowledge and social bias from large language models. We employ techniques based on computational learning theory and study which theoretical guarantees can be given for the correctness of the rules extracted.

RT9 TEAM LEADERS

- **Professor Anna Smajdor**, UiO, Department of Philosophy, Classics, History of Art and Ideas

- **Professor Manuela Zucknick**, UiO, OCBE

RT9 TEAM

- **Associate Professor Ana Ozaki**, UiO, Department of Informatics

- **Associate Professor Andrei Kutuzov**, UiO, Department of Informatics

- **Associate Professor Anne-Marie George**, UiO, Department of Informatics

- **Professor Arnaldo Frigessi**, UiO, Department of Mathematics/OCBE

- **Postdoctoral Research Fellow Camilla Lingjærde**, UiO, Department of Mathematics

- **PhD Fellow Emil Poiesz**, UiO, Department of Informatics

- **Professor Erik Velldal**, UiO, Department of Informatics

- **PhD Fellow Hannah Clausen**, UiO, Department of Informatics

- **Associate Professor Ida Scheel**, UiO, Department of Mathematics

- **Associate Professor Johan Pensar**, UiO, Department of Mathematics

- **Professor Lilja Øvrelid**, UiO, Department of Informatics

- **Postdoctoral Fellow Nikolay Arefev**, UiO, Department of Informatics

- **Associate Professor Samia Touileb**, UiB Department of Information Science and Media Studies



*As a philosopher, I tend to think I have a pretty clear idea what is meant by terms such as bias, or privacy; **part of the uniqueness of Integreat** is to see how other disciplines work with these concepts sometimes in very different ways, and how we manage to communicate across these differences. The really exciting possibilities, to my mind, come from these intersections, but there's no denying they are extremely challenging!*

– PI Ethics, Anna Smajdor,
Department of Philosophy, Classics, History of Art and Ideas

Publications



[Visit webpage](#)

al Hajj, G., **Hubin, A.**, Kanduri, C., Pavlovic, M., Rand, K.D., Widrich, M., **Solberg, A.S.**, Greiff, V., **Pensar, J.**, Klambauer, G. & **Sandve, G.K.** (2024). Incorporating probabilistic domain knowledge into deep multiple instance learning. In: Proceedings of the *Forty-first International Conference on Machine Learning (ICML)*, 235, 17279-17297.

<https://proceedings.mlr.press/v235/hajj24a.html>

Akman, T., Arendt, L.M., Geisler, J., Kristensen, V.N., **Frigessi, A.** & **Köhn-Luque, A.**, 2024. Modeling of Mouse Experiments Suggests that Optimal Anti-Hormonal Treatment for Breast Cancer is Diet-Dependent. *Bulletin of Mathematical Biology*, 86(4), p.42.

<https://doi.org/10.1007/s11538-023-01253-1>

Belman, S., **Pesonen, H.**, Croucher, N.J., Bentley, S.D. & **Corander, J.** (2024). Estimating between-country migration in pneumococcal populations. *G3 Genes[Genomes]Genetics*, 14(6).

<https://doi.org/10.1093/g3journal/jkae058>

Blomstedt, P., Mesquita, D., Rivasplata, O., Lintusaari, J., Sivula, T., **Corander, J.** & Kaski, S. (2024) Meta-Analysis of Bayesian Analyses. *Bayesian Analysis*, 1-26.

<https://doi.org/10.1214/24-BA1465>

Brandsæter, A. & **Glad, I.K.**, 2024. Shapley values for cluster importance: How clusters of the training data affect a prediction. *Data Mining and Knowledge Discovery*, 38(5), pp.2633-2664.

<https://doi.org/10.1007/s10618-022-00896-3>

Brandsæter, A. & **Glad, I.K.**, 2025. XAI in hindsight: Shapley values for explaining prediction accuracy. *Expert Systems with Applications*, p.126845.

<https://doi.org/10.1016/j.eswa.2025.126845>

Chan, Y.H., Rø, G., Midtbø, J.E., Di Ruscio, F.; Watle, S.S.V., Juvet, L.K., Littmann, J., Aavitsland, P., Nygård, K.M., Berg, A.S., Bukholm, G., Kristoffersen., A.B., Engø-Monsen, K., Engebretsen, S., Swanson, D., **Diz-Lois Palomares, A.**, Lindstrøm, J.C., **Frigessi A.** & de Blasio, B.F. (2024). Modeling geographic vaccination strategies for COVID-19 in Norway. *PLoS Computational Biology*, 20(1).

<https://doi.org/10.1371/journal.pcbi.1011426>

de Aguas Pérez, J.M., **Pensar, J.** & Biele, G. (2024). Partial identification and efficient estimation for the stratum-specific probability of benefit with thresholds on a continuous outcome. In: 9th Causal Inference Workshop at UAI, Barcelona, Spain.

<https://openreview.net/forum?id=JJH3ygg9z6>

de Aguas Pérez, J.M., **Pensar, J.**, Varnet Pérez, T.F. & Biele, G. (2024). *Recovering target causal effects from post-exposure selection induced by missing outcome data*. In: The 2024 American Causal Inference Conference, Seattle, USA.

<https://app.cristin.no/results/show.jsf?id=2307164>

George, A.-M. & **Dimitrakakis, C.** (2024). Eliciting Kemeny Rankings. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(11), 12164-12171.

<https://doi.org/10.1609/aaai.v38i11.29105>

Gkolemis, V., Gutmann, M. & **Pesonen, H.** (2024). An Extendable Python Implementation of Robust Optimization Monte Carlo. *Journal of Statistical Software*, 110(2), 1-26.

<https://doi.org/10.18637/jss.v110.i02>.

Griesbauer, E., Czado, C., **Frigessi, A.** & **Hobæk Haff, I.** (2025). TVineSynth: A Truncated C-Vine Copula Generator of Synthetic Tabular Data to Balance Privacy and Utility. In: *28th International Conference on Artificial Intelligence and Statistics (AISTATS 2025)*.

<https://doi.org/10.48550/arXiv.2503.15972>

Hjort, A.D., **Scheel, I.**, Sommervoll, D.E. & **Pensar, J.** (2023). Locally interpretable tree boosting: An application to house price prediction. *Decision Support Systems*, 178, 114106.

<https://doi.org/10.1016/j.dss.2023.114106>

Hjort, A.D., Williams, J.P. & **Pensar, J.** (2024). Clustered Conformal Prediction for the Housing Market. In: *Proceedings of 13th Symposium on Conformal and Probabilistic Prediction with Applications*, PMLR, 230, 366-386.

<https://proceedings.mlr.press/v230/hjort24a.html>

Hubin, A. & **Storvik, G.O.** (2024). Sparse Bayesian Neural Networks: Bridging Model and Parameter Uncertainty through Scalable Variational Inference. *Mathematics*, 12(6), 788.

<https://doi.org/10.3390/math12060788>

Jullum, M., **Aas, K.**, **Løland, A.**, Aase, F. S. & Olsen, L.H.B. (2024). *On conditional Shapley values for prediction explanation - Adaptive & variance stabilizing estimation with KernelSHAP*. In: InterACT workshop 2, Bremen, Germany.

<https://app.cristin.no/results/show.jsf?id=2305843>

Järvenpää, M. & **Corander, J.** (2024). Approximate Bayesian inference from noisy likelihoods with Gaussian process emulated MCMC. *JMLR*, 25(366), 1-55.

<https://www.jmlr.org/papers/v25/21-0421.html>

Kindermann, C., **George, A.-M.**, Parsia, B. & Sattler, U. (2024). Minimal Macro-Based Rewritings of Formal Languages: Theory and Applications in Ontology Engineering (and Beyond). In: *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(9), 10581-10588.

<https://doi.org/10.1609/aaai.v38i9.28928>

Lingjærde, C., Fairfax, B.P., **Richardson, S.** & Ruffieux, H. (2024). Scalable multiple network inference with the joint graphical horseshoe. *The Annals of Applied Statistics*, 18(3), 1899-1923.

<https://doi.org/10.1214/23-AOAS1863>

Mikkola, P., Martin, O.A., Chandramouli, S., Hartmann, M., Pla, O.A., Thomas, O.M.T., **Pesonen, H.**, **Corander, J.**, Vehtari, A., Kaski, S., Bürkner, P.-C. & Klami, A. (2024). Prior Knowledge Elicitation: The Past, Present, and Future. *Bayesian Analysis*, 19(4), 1129-1161.

<https://doi.org/10.48550/arXiv.2112.01380>

Moen, P.A.J., **Glad, I.K.** & **Tveten, M.** (2024). Efficient sparsity adaptive changepoint estimation. *Electronic Journal of Statistics*, 18(2), 3975-4038.

<https://doi.org/10.1214/24-EJS2294>

Olsen, L.H.B., **Glad, I.K.**, **Jullum, M.** & **Aas, K.** (2024). A comparative study of methods for estimating model-agnostic Shapley value explanations. *Data mining and knowledge discovery*, 38, 1782-1829.

<https://doi.org/10.1007/s10618-024-01016-z>

Pavlović, M., al Hajj, G., Kanduri, C., **Pensar, J.**, Wood, M. E., Sollid, L. M., Greiff, V. & **Sandve G.K.** (2024). Improving generalization of machine learning-identified biomarkers using causal modelling with examples from immune receptor diagnostics. *Nature Machine Intelligence*, 6(1), 15-24.

<https://doi.org/10.48550/arXiv.2204.09291>

Pesonen, H. & **Corander, J.** (2025). Stratified distance space improves the efficiency of sequential samplers for approximate Bayesian computation. *Computational Statistics & Data Analysis*, 207.

<https://doi.org/10.1016/j.csda.2025.108141>

Pflueger, M., Marti, J. & **Kostylev, E.V.** (2024). A characterisation theorem for two-way bisimulation-invariant monadic least fixpoint logic over finite structures. In: *Proceedings of the 39th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS 2024)*, 63, 1-14.

<https://doi.org/10.1145/3661814.3662107>

Pflueger, M., Tena Cucala, D.J. & **Kostylev, E.V.** (2024). Recurrent graph neural networks and their connections to bisimulation and logic. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(13), 14608-14616.

<https://doi.org/10.1609/aaai.v38i13.29377>

Pöntinen, A.K., Gladstone, R.A., **Pesonen, H.**, Pesonen, M., Cléon, F., Parcell, B.J., Kallonen, T., Simonsen, G.S., Croucher, N.J., McNally, A., Parkhill, J., Johnsen, P.J., Samuelsen, Ø. & **Corander, J.** (2024). Modulation of multidrug-resistant clone success in Escherichia coli populations: a longitudinal, multi-country, genomic and antibiotic usage cohort study. *The Lancet Microbe*, 5(2), e142-e150.

[https://doi.org/10.1016/s2666-5247\(23\)00292-6](https://doi.org/10.1016/s2666-5247(23)00292-6)

Ramezani-Kebrya, A., Antonakopoulos, K., Cevher, V., Khisti, A. & Liang, B. (2024). On the Generalization of Stochastic Gradient Descent with Momentum. *Journal of Machine Learning Research (JMLR)*, 25, 1-56.

<https://www.jmlr.org/papers/v25/22-0068.html>

Redelmeier, A.A., **Jullum, M., Aas, K. & Løland, A.** (2024). MCCE: Monte Carlo sampling of valid and realistic counterfactual explanations for tabular data. *Data mining and knowledge discovery*, 1830-1861.

<https://doi.org/10.1007/s10618-024-01017-y>

Schmiester, L., Brasó-Maristany, F., González-Farré, B., Pascual, T., Gavilá, J., Tekpli, X., Geisler, J., Kristensen, V.N., **Frigessi, A.**, Prat, A. & **Köhn-Luque, A.** (2024). Computational Model Predicts Patient Outcomes in Luminal B Breast Cancer Treated with Endocrine Therapy and CDK4/6 Inhibition. *Clinical Cancer Research*, 30(17), 3779-3787.

<https://doi.org/10.1158/1078-0432.ccr-24-0244>

Segal, M., **George A.-M.**, Chieh Yu, I. & **Dimitrakakis, C.** (2024). Better Luck Next Time: About Robust Recourse in Binary Allocation Problems. In: *World Conference on Explainable Artificial Intelligence*, pp. 374-394, Springer Nature Switzerland.

https://doi.org/10.1007/978-3-031-63800-8_19

Segal, M., **George, A.-M. & Dimitrakakis, C.** (2024). *Threshold Recourse for Dynamic Allocations*. In Third European Workshop on Algorithmic Fairness, Mainz, Germany.

<https://app.cristin.no/results/show.jsf?id=2355528>

Skaaret-Lund, L., **Storvik, G.O. & Hubin, A.** (2024). Sparsifying Bayesian neural networks with latent binary variables and normalizing flows. *Transactions on Machine Learning Research (TMLR)*, 10, 1-32.

<https://doi.org/10.48550/arXiv.2305.03395>

Štrumbelj, E., Bouchard-Côté, A., **Corander, J.**, Gelman, A., Rue, H., Murray, L., **Pesonen, H.**, Plummer, M. & Vehtari, A. (2024). Past, Present and Future of Software for Bayesian Inference. *Statistical Science*, 39(1), 46-61.

<https://doi.org/10.1214/23-ST907>

Varnet Pérez, T.F., Øvergaard, K. R., **Frigessi, A.** & Biele, G. (2024). *Design-aware imputation in a target trial of ADHD pharmaceutical treatment on Norwegian national test scores: the case of simultaneous outcomes*. In: European Causal Inference Meeting (EuroCIM), Copenhagen, Denmark.

<https://app.cristin.no/results/show.jsf?id=2307099>

Walega, P.A., Tena Cucala, D.J., Cuenca Grau, B. & **Kostylev, E.V.** (2024). The Stable Model Semantics of Datalog with Metric Temporal Operators. *Theory and Practice of Logic Programming*, 24(1), 1-35.

<https://doi.org/10.1017/S1471068423000315>

Wold, S., Simon, É., Charpentier, L.G.G., **Kostylev, E.V., Velldal, E. & Øvrelid, L.** (2024). Compositional Generalization with Grounded Language Models. In: *Findings of the Association for Computational Linguistics ACL 2024*, pp. 3447-3460. Bangkok, Thailand: Association for Computational Linguistics.

<https://doi.org/10.18653/v1/2024.findings-acl.205>

Wu, C., Gunnarsson, E.B., Myklebust, E.M., **Köhn-Luque, A.**, Tadele, D. S., Enserink, J., **Frigessi, A.**, Foo, J. & Leder K. (2024). Using birth-death processes to infer tumor subpopulation structure from live-cell imaging drug screening data. *PLoS Computational Biology*, 20(3).

<https://doi.org/10.1371/journal.pcbi.1011888>

Wu, Z., Choi, C., Cao, X., Cevher, V. & **Ramezani-Kebrya, A.** (2025). Addressing Label Shift in Distributed Learning via Entropy Regularization. In: *Proceedings of the International Conference on Learning Representations (ICLR)*, 2024.

<https://doi.org/10.48550/arXiv.2502.02544>

Xie, W., Pethick, T., **Ramezani-Kebrya, A.** & Cevher, V. (2024). Mixed Nash for Robust Federated Learning. *Transactions on Machine Learning Research*, 2024.

<https://infoscience.epfl.ch/entities/publication/d6c4f9c0-be2b-4949-ae7-7c748d9617c2>

Athanasopoulos, A., **George, A.-M. & Dimitrakakis, C.** (2025). Probably Correct Optimal Stable Matching for Two-Sided Markets Under Uncertainty. In: *Proceedings of 24th International Conference on Autonomous Agents and Multiagent Systems*, 2025.

<https://doi.org/10.48550/arXiv.2501.03018>

de Aguas, J., **Pensar, J.**, Varnet Pérez, T., & Biele, G. (2025). Recovery and inference of causal effects with sequential adjustment for confounding and attrition. *Journal of Causal Inference*.

<https://doi.org/10.48550/arXiv.2401.16990>

Thomas, O., Sá-Leão, R., de Lencastre, H., Kaski, S., **Corander, J. & Pesonen, H.** (2024). Misspecification-robust likelihood-free inference in high dimensions. *Computational Statistics*, 2025.

<https://doi.org/10.48550/arXiv.2002.09377>

Eide, S. & **Frigessi, A.** (2025). BoRA: Bayesian Hierarchical Low-Rank Adaption for Multi-Task Large Language Models. In: *Northern Lights Deep Learning Conference 2025*, 265, 51-57.

<https://proceedings.mlr.press/v265/eide25a.html>

Preprints:

George, A.-M. (2024). Hypergeometric Distribution Revisited: Tail Inequalities, Confidence Bounds and Sample Sizes. *ARxiv*.

<https://doi.org/10.48550/arXiv.2405.06722>

Järvenpää, M., **Corander, J. & Pesonen, H.** (2025). Surrogate-based ABC matches generalized Bayesian inference under specific discrepancy and kernel choices. *ARxiv*.

<https://doi.org/10.48550/arXiv.2502.11738>

Olsen, L.H.B. & Jullum, M. (2024). Improving the Sampling Strategy in KernelSHAP. *arXiv: arxiv:2410.04883*. doi: 10.48550/arXiv.2410.04883. aRxiv.

<https://doi.org/10.48550/arXiv.2410.04883>

Jullum, M., Olsen, L.H.B., Lachmann, J. & Redelmeier, A. (2024). shapr: Explaining Machine Learning Models with conditional Shapley values in R and Python”. *R-Project*.

https://cran.r-project.org/web/packages/shapr/vignettes/general_usage.html

Ojala, F., **Pesonen, H.**, Gladstone, R.A., Mäklin, T., Tonkin-Hill, G., Marttinen, P. & **Corander, J.** (2025). Basic reproduction number for pandemic Escherichia coli clones varies markedly and can be comparable to pandemic influenza viruses. *bioRxiv*.

<https://doi.org/10.1101/2024.05.08.593267>

Moreira-Soares, M., Fossen, E.I.F., Bilbao-Jayo, A., Almeida, A., Lopez-Perez, L., Alonso, I., Cabrera-Umpierrez, M.F., Fico, G., Singer, S., Taylor, K.J., Ness, A., Thomas, S., Pring, M., Licitra, L., Cavalieri, S., **Frigessi, A.** & LeBlanc, M. (2024). Joint probability approach for prognostic prediction of conditional outcomes: application to quality of life in head and neck cancer survivors. *medRxiv*.

<https://doi.org/10.1101/2024.12.16.24319067>

Azimzade, Y., Haugen, M.H., Kristensen, V.N., **Frigessi, A. & Köhn-Luque, A.** (2024). Integrated multiomics analysis unveils how macrophages drive immune suppression in breast tumors and affect clinical outcomes. *BioRxiv*.

<https://doi.org/10.1101/2024.11.09.622776>

Myklebust, E.M., **Frigessi, A.**, Schjesvold, F., Foo, J., Leder, K. & **Köhn-Luque, A.** (2024). Prediction of cancer dynamics under treatment using Bayesian neural networks: A simulated study. *arXiv*.

<https://doi.org/10.48550/arXiv.2405.14508>

Günther, F., Brustad, H.K., **Frigessi, A.** & Britton, T. (2024). Quantifying the impact of social activities on SARS-CoV-2 transmission using Google mobility reports. *MedRxiv*.

<https://doi.org/10.1101/2024.01.03.24300755>

Eliseussen, E., **Frigessi, A. & Vitelli, V.** (2023). Rank-based Bayesian clustering via covariate-informed Mallows mixtures. *arXiv*.

<https://doi.org/10.48550/arXiv.2312.12966>

D’Alessandro, M., Asenso, T.Q. & **Zucknick, M.** (2024). Integrating Multiple Data Sources with Interactions in Multi-Omics Using Cooperative Learning. *arXiv*.

<https://doi.org/10.48550/arXiv.2409.07125>

International collaborations

By welcoming researchers from around the world, we create a dynamic environment where cutting-edge interdisciplinary research in ML can flourish. Fostering a space where both experts and early-career scholars exchange ideas and build professional networks is crucial for advancing research and tackling complex scientific challenges.

In 2024, we invited experts to inspire and contribute to our research projects, with their diverse perspectives and methodologies, laying the foundation for collaboration. Our visiting fellowships allow researchers from different countries to spend time at Integreat. In turn, our researchers make exchange visits to exciting AI groups. We invite international researchers to give guest lectures. We have also hosted for longer periods master and PhD students from our international collaborations, supervising their research within Integreat's projects.

We were pleased to welcome these distinguished academic visitors at Integreat:

- **Volkan Cevher**, Associate Professor of Mathematics at EPFL, working on optimisation and knowledge-driven ML.
- **Claudia Czado**, Professor of Applied Mathematical Statistics at the Technical University of Munich, collaborating with us in modelling complex dependence and vine copula in ML.
- **Idris Eckley**, Professor of Statistics and Director of the centre STORi in Lancaster University, whose research in statistical methods and data science for anomaly and changepoint detection is highly relevant to our work.
- **Rebecka Jørnsten**, Professor of Biostatistics and Applied Statistics at the University of Gothenburg/Chalmers, with focus on biostatistics and statistical ML methodology in health-related problems and predictive modelling.
- **Eric Laber**, Professor of Statistical Science and Co-Director of the Duke Computing Initiative at Duke University, presenting on causal inference and ML.

We welcomed at Integreat these early-career researchers:

- **Sara Spieler**, Master student in Bioinformatics, Université Côte d'Azur, Nice, France. Integreat research project Hybrid models for personalised medicine, modelling drug response in breast cancer using high-dimensional molecular data. 6.01.- 30.06.2025
- **Kristina Bratkova**, PhD student, Lancaster University, UK 10.03-21.03.2025
- **Ana Niño López**, PhD student in Mathematics, Universidad de Cádiz, Spain. Integreat research project Hybrid models for personalised medicine, modelling regeneration of immune cell populations in childhood leukemia after treatment using bone marrow and blood tests. 17.03.- 13.06.2025

Integreat has started to build special partnerships with several other centres around the world. This currently includes:

- The **Prob_AI Hub** is a research hub funded by EPSRC in the UK to develop better understanding of and new methods in AI. This will be achieved by research at the interface of AI and Mathematics. The Prob_AI hub will focus on probabilistic AI and bring researchers with skills across areas such as Bayesian and Computational Statistics, Dynamical Systems, Numerical Analysis, PDEs, Probability, Stochastic Analysis, and Uncertainty Quantification to work with AI and Machine Learning researchers. Prob_AI includes the universities of Bristol, Cambridge, Edinburgh, Lancaster, Manchester. and Warwick. Frigessi chairs the Advisory Board.
- **STOR-i** Centre for Doctoral Training is a joint centre between the School of Mathematical Sciences and the Management Science department of Lancaster University. Integreat's PIs supervise three PhD students at STORi, who spend short periods in Oslo.. Glad is a member of the Scientific Advisory Committee of STORi, and Idris Eckley, director of STORi, is visiting professor at Integreat.
- The Medical Research Council (MRC) **Biostatistics Unit** (BSU) at the University of Cambridge is one of the largest groups of biostatisticians in Europe, and a major centre for research, training and knowledge transfer in biostatistics and biostatistical ML. Integreat and the BSU are preparing a memorandum of collaboration which will open for new collaborations across the two centres. Sylvia Richardson, former director of the BSU and professor emerita at the University of Cambridge, is visiting professor at Integreat.
- Integreat's PI have a close collaboration with the **Universities of Hawassa and Jimma** in Ethiopia, where they supervise PhD students and contribute to teaching PhD courses. The collaboration has been ongoing for more than ten years.

We are working on new centre-to-centre agreements.



« *Visiting Integreat in person in 2024 was a great experience. The working environment is excellent and I appreciated especially the open atmosphere for interaction.* »

– Professor Claudia Czado, Technical University of Munich, visiting researcher 2024.



« *Over the last year, I have thoroughly enjoyed Integreat's inclusive environment, sharing my statistical expertise and expanding my knowledge into AI thanks to many stimulating project interactions with machine learners and computer scientists.* »

– Sylvia Richardson, Professor University of Cambridge, MRC Biostatistics Unit, OCBE

Integreat for learning and collaboration: meetings, reading clubs, workshops, and seminars

Integreat hosts research groups from different disciplines: statistics, logic, language technology, theoretical computer science, mathematics, and, of course, ML.

We have regular activities fostering learning, collaboration, and critical engagement across these research groups. Activities include meetings, reading clubs, seminars, and workshops. They provide dedicated arenas for discussing emerging ideas and methodologies and engaging with theoretical and practical challenges. These groups are dynamic and change in line with the academic progress, demands and priorities. These are forums where participants encounter diverse perspectives that challenge assumptions and push the boundaries of their fields. Integreat’s annual meetings gather all researchers and students affiliated with the centre.

Regular seminars and meetings

- **Integreat Tuesday seminars**
Public seminar series, open to all interested. The Tuesdays seminars are devoted to themes that are relevant to Integreat. Every semester, half of the Tuesday seminars are dedicated to one of the four objectives of Integreat. In 2024 this was explainable ML and sustainable ML, in 2025 we concentrate on uncertainty quantification first and then on accuracy of algorithms, with perspectives related to alignment and generalisability. Seminars last for 45 minutes and are followed by questions and discussion. All seminars are streamed. Subscribe to our open events <https://sympa.uio.no/integreat.no/subscribe/seminars>
- **Integreat Wednesday General Assembly**
On Wednesdays, the Integreat community gathers for a short internal seminar, followed by discussions and communication. A simple lunch is served. The seminar covers topics such as recently submitted papers, conference reports, and introductions of new Integreat members and visitors. The meeting is streamed between Oslo and Tromsø.
- **Thursday Tromsø-Oslo coffee**
A 30-minute weekly informal chat across the two hubs.

Reading groups

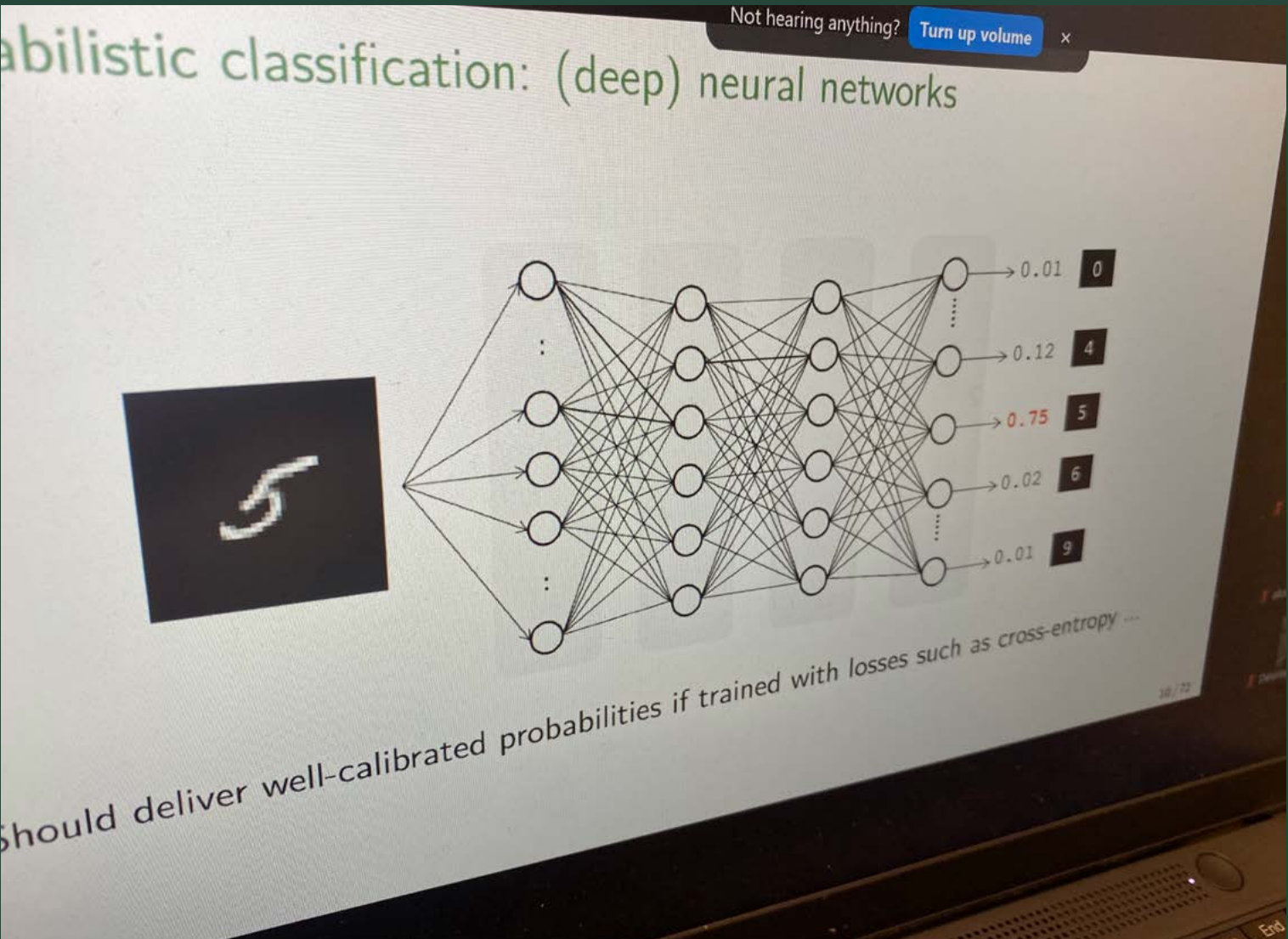
Reading groups start spontaneously and are dynamic. Some are open to everybody interested, others are by invitation only as they are linked to specific Integreat projects. These groups have been or are currently active:

- **Discrepancies measures working group**
- **Hybrid models - Data Augmentation in ML**
- **Learning from human preferences**
- **Inconsistency learning club**
- **Reading group on Bayesian neural networks**
- **Reading Group on data-driven tensor factorisation models**

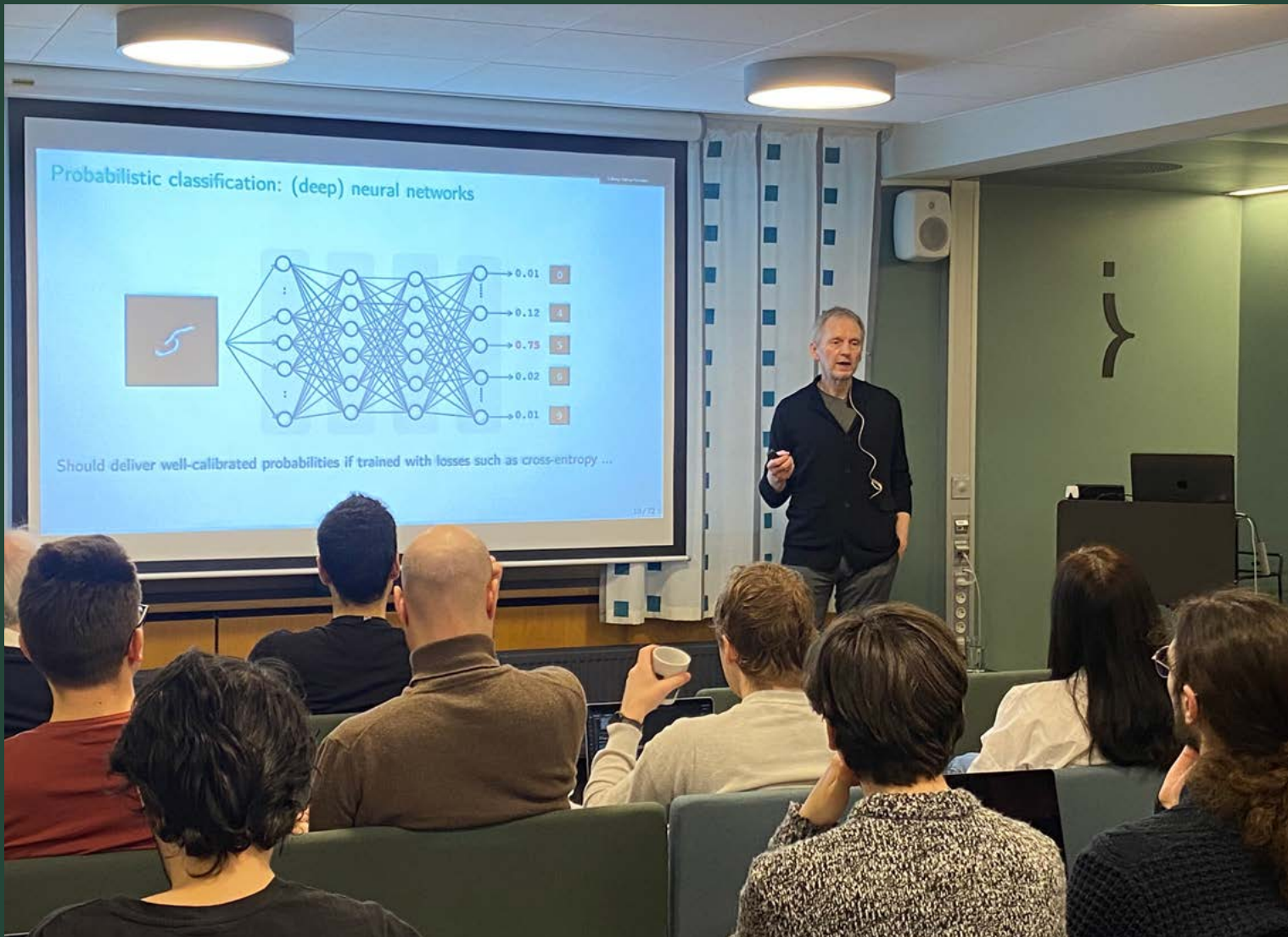
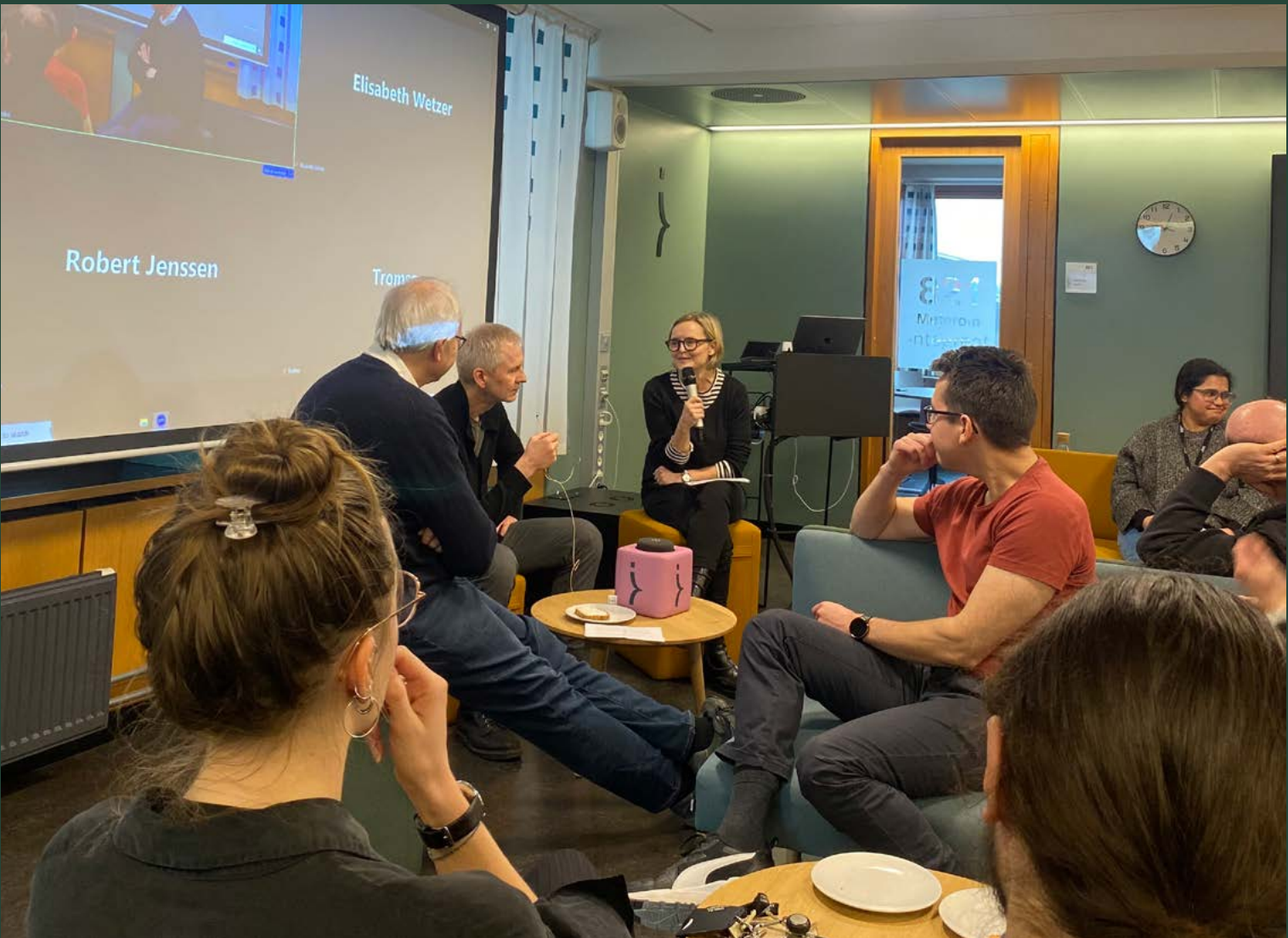
Thematic research seminars

Some of Integreat’s research groups run their own seminars:

- **Visual intelligence seminars**
- **Tuesday Seminar in Statistics and Data Science**
- **Biostatistical seminars**



Co-director Professor Lilja Øvrelid presents Aksel Braanen Sterri



Eyke Hüllermeier presenting at Integreat, February 2025



Integreat Tuesday seminar: Aksel Braanen Sterri

From the Tutorial on Uncertainty in Machine Learning with Professor Eyke Hüllermeier



Research can sometimes feel like living in a single-room house—no matter how well you know it, there’s always more beyond its walls. The seminars in our group are like doors opening into new rooms, offering perspectives from diverse fields that challenge and expand our thinking. In a time when being a ‘jack of all trades’ is just as valuable as deep expertise, these sessions encourage us to explore beyond our immediate focus, making connections that enrich our work in unexpected ways. I truly appreciate the effort that goes into organising them—they don’t just add to our knowledge; they broaden the way we see it.

– PhD Fellow Kushagri Tandon, Department of Mathematics

Integreat events

Integreat's launch was not a single event, but a series of happenings spread throughout September 2023 to the end of 2024. From the welcoming housewarming in Oslo to the inaugural kick-off, from the official opening with the Minister of Higher Education and Research to the first annual retreat in Tromsø, these moments brought together the Integreat community, fostering collaboration and new friendships across institutions and disciplines.



Norwegian Minister of Digitalisation and Public Governance Karianne Tung held the opening lecture on digitalisation status in Norway on 31 October 2024.

Our New Home: The Oslo Hub Housewarming

In October 2023, Integreat opened its doors to the newly renovated eighth floor of the Niels Henrik Abels building in UiO's Blindern campus. The Department of Mathematics has generously provided offices and workspaces, meeting rooms, and a common lounge for Integreat. The renovation was inspired by the original architecture and internal design of the 1965 Abels hus, with green and yellow colours on the walls, refurnished spaces, glass doors, and a new North-West-South-East axis. The floor is shared with the Section of Statistics and Data Science of the Department of Mathematics. The housewarming was a celebration of the collective effort that went into creating the new home for the Centre of Excellence. Researchers, staff, and collaborators gathered to mark the occasion, enjoying good company and the best Italian ice cream available in Norway, and served personally by Centre Director Arnoldo Frigessi. The leadership team expressed their gratitude to everyone who contributed—the architects and estate planners, the workers and the administration of the department that welcomed Integreat into their space.



Building Momentum: The Integreat Kick-off

With the Oslo hub officially opened, Integreat moved into its next phase with a kick-off event in 2023. This gathering brought together researchers to discuss the centre's vision and research plans. It was a chance to align efforts and set the stage for impactful collaborations, reinforcing Integreat's commitment to interdisciplinary research and ethical AI development. The plans for first research projects and their teams were presented.



The Official Launch: A Landmark Event



[Visit webpage](#)

The official launch of Integreat took place at University of Oslo in August 2024. After the welcome by UiO Rector Svein Stølen, the Minister of Research and Higher Education Oddmund Hoel gave his speech, focusing on AI in Norway and the expected role of Integreat. The CEO of the Research Council of Norway Mari Sundli Tveit then presented the SFF plaque to Frigessi, congratulating for the centre. The centre director gave his thanks and recognised the responsibilities that Integreat has. Thereafter, there were greetings by the other two partners of Integreat, André Teigland, CEO of the Norwegian Computing Center, and Olav Gaute Hellestø, head of Department of Physics and Technology at UiT. Next, six members of Integreat described what their dreams about Integreat are: Robert Jenssen, Lilja Øvrelid, Anne-Marie George, Martin Giese, Ingrid Glad, and Vilde Ung. A panel debate with all authorities that were present concluded the event.



Coming Together: The First Annual Retreat

The first Integreat Annual Retreat was held at UiT in Tromsø, November 25–27, 2024, bringing together the entire Integreat family—researchers, PhD fellows, postdocs, and staff.

Held in the inspiring Arctic setting, the retreat was an opportunity to strengthen connections and exchange ideas across projects. The programme included the presentation of all projects running in or around Integreat. We had an unforgettable evening excursion to the Wild Nature Centre, with the Northern lights in the background. Each Research Theme summarised its status and plans.

As the centre moves forward, the collaborations and connections formed during these first events will develop and bloom. With an engaged and growing network of researchers, Integreat is poised to reach its ambitious aims.





Integreat meets the public

Science at Integreat is open. We adhere to Plan-S (<https://www.coalition-s.org/>) for open access publishing. Computer codes are open- source and made available online and documented. All our results will be published in world-leading scientific journals and peer-reviewed conferences. Our webpages present material for specialists and for the general public. We contribute to the annual UiO Data Science Day as a meeting point for data scientists from academia, industry, and public sector. We will present at and organise international workshops and conferences, and support the Northern Lights Deep Learning Conference.

Integreat is active in the social media and regularly present in the Norwegian media. We will arrange meetings to discuss ethical aspects of our work to combat biases and discrimination in and because of AI. Integreat is a clear voice in the public debate on AI, giving public and direct advice to the administration, and participating to national committees, with the aim to make Norway one of the best places in the world to access and use safe, secure and high-quality ML.

For more information about the dissemination work of Integreat, and our presence in the media, please see our [webpages](#).



[Visit webpage](#)

Integreat in media



Publisert 19. feb. 2025 11:18

Kunstig intelligens: – Algoritmene skal jobbe for oss, ikke omvendt

Integreat på Forskning.no: Hvordan kan KI bli smartere, mer etisk og bedre tilpasset oss mennesker? Forskere mener vi står ovenfor en ny æra i hvordan algoritmer bygges.



Publisert 19. feb. 2025 10:29

Integreat i podcasten Nysgjerrige Norge

Lurer du på hva vi forsker på og prøver å få til her i Integreat? Da anbefaler vi deg å høre podcastepisoden Nysgjerrige Norge - Episode 3: Et dypdykk i kunnskapsdrevet maskinlæring med Integreat.



Publisert 20. nov. 2024 10:16

Det trengs bedre og sikrere KI-verktøy for forskning

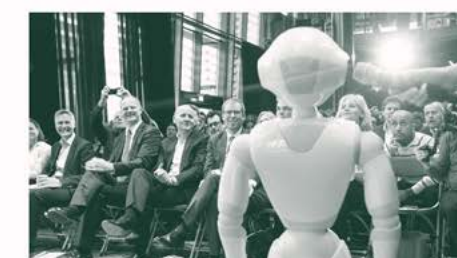
Integreat i Khrono: Dagens verktøy for kunstig intelligens er for dårlig på andre språk enn engelsk, er ikke tilpasset europeiske lover og praksis og er ikke tverrfaglige nok for forskningsformål, slår et nytt notat fra universitetsalliansen The Guild fast.



Publisert 1. nov. 2024 09:04

– Vi skal bygge og trygge Norge med digitalisering som verktøy

Minister Karianne Tung viste retningen videre for Norge under Digitaliseringsforelesningen ved Universitetet i Oslo.



Publisert 11. okt. 2024 10:13

Syv forslag til hvordan Norge kan bruke mer kunstig intelligens

Norge er i verdenstoppen i digitale tjenester, men henger etter med å implementere kunstig intelligens. Ledergruppen i Integreat har syv forslag til hvordan man kan legge til rette for mer bruk av kunstig intelligens i Norge og mer verdiskaping.

Norwegian political gathering – Integreat at Arendalsuka 2024

At Arendalsuka 2024 – the largest political gathering in Norway – Integreat hosted an event that focused on how Norway can increase the use of AI across sectors for the benefit of society. Ingrid Glad, Co-Director at Integreat, opened the event and introduced the centre. Arnaldo Frigessi, spoke with Guri Lande, Director of Digital Strategy and Coordination at the Norwegian Digitalisation Agency, about Norway's status and challenges related to AI. Lilja Øvrelid, also a Co-Director at Integreat, led a discussion with Cathrine Pihl Lyngstad, Head of Data at NAV, and Trygve Refvem, Head of the Innovation Team at the Norwegian Tax Administration. They discussed how AI can be implemented in the public sector, focusing on overcoming legal and bureaucratic challenges. Anders Løland, from the Norwegian Computing Center (NR), spoke with Eirik Andreassen, Head of Technology and Networks at DigitalNorway, and Anne-Grete Ellingsen, Entrepreneur for SME Financing, on how AI can become more accessible to SMEs through better training and research-near consulting. Ingrid Hobæk Haff, PI at Integreat, led a discussion with Terje Rootwelt, CEO of Helse Sør-Øst, and Eirik Vegler Broen, Tech Lead for Health Data at the City of Oslo. They explored how AI can improve patient care and streamline administrative processes in healthcare, focusing on the secure use of sensitive health data. We closed the event with a panel led by Ingrid Glad with Alfred Bjørlo (Member of Parliament, Venstre), Rune Støstad (Member of Parliament, Arbeiderpartiet), and Katrine Tjølsen (Head of MDG's AI Committee). Political measures to promote AI use in Norway were discussed, with an emphasis on public-private collaboration and the need to build competence across sectors. A major theme was data sharing and governance models, with a call for improved legal frameworks and cross-sector cooperation.



Integreat Digitalisation Lecture 2024

A new annual tradition launched by Integreat at the University of Oslo



[Visit webpage](#)

Integreat digitalisation lecture is an important annual meeting point between the minister of digitalisation and the researchers in Norway.. It is an opportunity for the Norwegian government to update the public on its strategies and plans for AI and it offers an arena where researchers and students can engage directly with the minister.

The 2024 Integreat Digitalisation Lecture, held at the University of Oslo on 31 October 2024, marked the beginning of the new annual tradition. The main speaker was the Norwegian Minister of Digitalisation and Public Governance (Digitaliserings- og forvaltningsminister) Karianne Tung. Based on the recently published National Digitalisation Strategy 2024–2030 “The digital Norway of the future”, the lecture of Minister Tung focused on the government’s ambition to make Norway one of the world’s most digitally advanced country by 2030, highlighting key areas such as the maritime industry, aquaculture, and health technology, where AI will provide significant benefits. **“Norway can become a world leader in the use of AI in these sectors, and we have the potential to**

lead the way in the safe and ethical use of AI,” Tung stated. For Norway, it is essential that AI is secure and follows the legislations and regulations, “One of the key areas where we can contribute is to facilitate data sharing,” she said, acknowledging that sharing data is essential for unlocking the full potential of AI. Minister Tung underlined the government’s support to research, including the billion-kroner fund for AI research. Tung reassured the audience that further steps would be taken to enhance the connection between ai research, public policy, and industry needs. She outlined plans to develop the national AI infrastructure, with high-performance computing capacity for research and public institutions. “We are committed to developing a Norwegian Large Language Model,” she

said. After the lecture, a discussion panel took place. Arnoldo Frigessi, the head of Integreat, emphasised that the Digitalisation Lecture aims to become an annual tradition. “Our goal is for the Digitalisation Lecture to be a long-standing tradition that brings together political leaders, academics, and industry experts to discuss and shape the future of digitalisation in Norway,” said Frigessi. Later, commenting on her lecture, Minister Tung wrote on LinkedIn “Being allowed to delve into a topic - and especially AI -, having the time to reflect on it and discuss it with students, academic and administrative staff is simply a joy. This way we can all become a little wiser and understand different perspectives.”







At Integreat, The Norwegian Minister of Research and Higher Education announced a major investment in research

On 17 February 2025, the newly appointed Minister of Research and Higher Education, Sigrun Aasland, visited the University of Oslo (UiO) and Integreat to announce a significant boost in research funding. The government allocated an additional half a billion NOK to groundbreaking research. **“We need to address major societal challenges in the coming years. This requires both developing and applying new knowledge. Norway has outstanding research environments with great opportunities and a significant responsibility in this regard,”** said Minister Aasland. The choice of Integreat as the venue for this announcement highlights our role as a hub for cutting-edge basic research.

Minister Aasland and her delegation from the Ministry of Education and Research, alongside Benedicte Løseth and Petter Helgesen from the Research Council of Norway (RCN), met with UiO Rector Svein Stølen, Pro-Rector Åse Gornitzka, Integreat Directors Arnoldo Frigessi and Ingrid Glad, postdoctoral researcher Leiv Rønneberg, early career researcher Anne-Marie George, and other invited guests. Media representatives were also present to cover the event. Following a welco-

me by Stølen and Glad, Minister Aasland officially announced the increased funding. Benedicte Løseth provided further comments on behalf of the Research Council. This led to an engaging roundtable discussion where Frigessi presented Integreat, Rønneberg shared his experience of being part of a centre like Integreat, while George expressed her excitement about receiving, together with an Integreat PI Valeria Vitelli, a FRIPRO grant that very day. The session ended with media interviews with the minister.

After the official announcement of the new important funding to basic research, Minister Aasland joined a group of Integreat's PhD Fellows, Postdocs, and researchers, who had a working group on the recently released Chinese new language model DeepSeek, which had made headlines worldwide after it topped download charts and caused US tech stocks to fall significantly. Håkon Austlid Tasken and August Arnstad were holding a presentation for the group, then the Minister arrived. She contributed with interesting questions and showed her curiosity for our work. Her engagement was highly appreciated by us all!



From the left Ingrid Glad, Arnoldo Frigessi, Svein Stølen, Sigrun Aasland, and Benedicte Løseth.

Building the Integreat innovative researcher training programme for PhD and Postdoc fellows

As Integreat enters its second year, the centre is thriving with a vibrant community of young researchers with diverse perspectives, cultures, and interests. Our fellows are working well on their projects, and we are excited to see their growth, engagement, and success. At Integreat, we emphasise collaboration, mentorship, diversity, and career development to foster the next generation of ML researchers.

Autumn 2024 marked a pivotal moment for Integreat. The first cohort of 12 PhD and Postdoctoral Fellows joined the centre. We succeeded to announce all positions together, even though the employment happened in different universities and departments. We presented many possible interdisciplinary projects, and applicants were asked to rank up to five preferred ones. The selection process was intense, and we succeeded to recruit a strong group of young scientists. Now, the fellows have offices at the Integreat premises at UiO or UiT, where they sit and work together alongside their supervisors and other experienced researchers.

The recruited fellows come from a diverse range of disciplines: statistics, machine learning, logic, language technology, physics, computer science, and mathematics — some directly from their master studies (in Bergen, Cambridge, Copenhagen, New Delhi, Oslo, Oxford, Torino, Trento, or Trondheim, among others), others after a short experience in the industry (Norges bank, Folkehelsinstituttet, Banca Intesa, Orderfox AG, etc.). Currently Integreat has 25 PhD and Postdoctoral Fellows in total.

We must look for the important questions and work well together. Our PhD candidates have both ambition and technical competence to develop their ideas into solid results, and our research teams are here to guide them in achieving these goals.

– Centre Director Arnoldo Frigessi



It's a very social environment which I very much enjoy. I feel like it's a nice little community outside of the professional work environment as well. We can build friendships outside of work which I think is really nice and important.

– Integreat PhD Hannah Clausen, Department of Informatics, University of Oslo

Strengthening research and mentorship

Integreat is committed to the professional development of its early-career researchers, offering extensive support through mentorship, skill development, and career progression programmes. INTREF - Integreat young researchers' forum is the Integreat arena dedicated to these themes. The INTREF initiative is structured around three pillars: science, social skills, and transferable skills, ensuring that our fellows are prepared for diverse career paths in all sectors, whether it is academia, public, or the private sector, in Norway and internationally.

INTREF, owned and driven by the Fellows themselves and with support from leadership and administration, provides a range of activities covering all three pillars at different stages. One new instrument is what we call “research kitchens” - pop-up, small workshops. Research kitchens is an innovative, collaborative environment where researchers from diverse fields come together to “cook up” new ideas, solutions, or knowledge. Just as a

chef uses a set of ingredients to create something new and useful in the kitchen, researchers bring together various inputs — such as data, theories, methodologies, and expertise — to generate novel insights and breakthroughs.

In a research kitchen, the process is highly interactive and dynamic. Researchers meet in a shared space where they can collaborate, exchange ideas, experiment, and engage in discussions, much like cooking together in a kitchen. We believe that the act of working together in this “kitchen” fosters creativity, innovation, and interdisciplinary collaboration. The social aspect is, of course, also important.

The mentorship programme is one of the key components of the training programme, both in a traditional way, where we pair early career researchers with dedicated mentors, and in the reversed way – namely, the fellows take on the role of mentors, guiding the conversation with the senior researchers. This approach allows both



As newly hired Postdoctoral Fellows, it has been wonderful getting to know the vibrant research community at Integreat, and be able to take part in the groundbreaking research within the centre. Integreat has a stimulating forward momentum, with a large cohort of PhD students starting at the same time and strong projects sitting right at the cutting edge of current methodology.

– dSTrain Postdoctoral Fellows Camilla Lingjærde and Leiv Rønneberg, Department of Mathematics, University of Oslo

parties to benefit, with senior researchers gaining fresh insights and perspectives, while the younger fellows develop leadership, communication, and pedagogical skills.

Building a strong community

A strong sense of community is at the heart of Integreat. In the first few months of the programme, the fellows have already organised social events such as volleyball games, crocheting gatherings, and reading groups.

Associate professor Ingrid Hobæk Haff is responsible for INTREF from the side of the centre's leadership. She highlights the importance of cultivating a friendly and academically enriching atmosphere. Integreat provides an opportunity for collaboration between teams, supervisors, and peers, and we are working to further develop the career programme to ensure that our fellows have the support they need to succeed," says Ingrid.

Centre Director Frigessi adds, "It's wonderful to see how the students are helping each other. They've formed

study groups, where they post-process together lectures they have followed. Indeed, they are learning better than anyone else!"

The path forward: key initiatives for 2025 and beyond

Looking to the future, Integreat is committed to expanding and enhancing the opportunities available to our early-career researchers. The INTREF - Integreat young researchers' forum will be further developed in 2025, with initiatives designed to provide greater research autonomy, leadership skills, and networking opportunities. We will also focus on increasing the visibility and impact of our fellows' research through interdisciplinary collaborations and international research visits.

Our centre will also continue to prioritise well-being and work-life balance by offering flexible working hours, mental health support services, and discussions on stress management. Recognising and celebrating our fellows' achievements will be a priority, and we will honour milestones such as publications, grants received, awards, PhD completions, etc.



Inclusive excellence:

a commitment to equity,
diversity, inclusion,
and belonging



I remember from the beginning of my career that sometimes you feel the struggle to be seen, be considered worthy, be someone who has a voice in the conversation. And as a student, it was difficult at times. For example, you wouldn't so easily encounter a female person in leading positions, and therefore, you start wondering: is there any place for me really here? In the context of Integreat, I see that we are explicitly trying to address this: there are lots of female PIs, and we have a mixture of backgrounds, cultures, ways of working, and, of course, also gender. All this is well represented in Integreat community.

– Integreat PI, Associate Professor Valeria Vitelli, Institute of Basic Medical Sciences

At Integreat, we believe that diversity and inclusion stand as foundational principles, essential to creating an environment where scientific innovation, creativity, and collaboration bloom. These values guide every aspect of our organisation - from recruitment and career development to fairness and ethical alignment of our research and the inclusion of diverse voices in all our endeavours. In the first year of Integreat, we have focused on establishing these core values for inclusion, and it is our wish to ensure that our commitment is properly embedded in our work, fostering an environment where all individuals are encouraged to contribute their unique perspectives and talents. We believe that true diversity transcends gender and includes a variety of identities, beliefs, ethnicities, and life experiences. Integreat wants to contribute to democracy and citizen participation.

Progressing toward gender equality and representation

In 2024, Integreat established the general plan for its DEI (Diversity, Equity, and Inclusion) initiatives, setting the scene for future work and guiding principles that will shape its impact in the years ahead. We strive to achieve gender balance within researcher teams, the leadership group, and administration. By focusing on an inclusive policy, we aim to ensure equal visibility for both female and male staff at all career levels, making Integreat a welcoming and attractive workplace for everyone.

While we have specifically targeted efforts to recruit female candidates for new positions, we have not entirely succeeded yet. There is still much to be done, as reflected in the recruitment of only four female candidates in 14 positions. The fields of computational sciences, both in Norway and internationally, remain predominantly

male, and Integreat is dedicated to continuing its efforts to foster greater diversity.

Integreat PhD fellow, Hannah Clausen, exemplifies the power of diversity in driving transformative research. Hannah's work focuses on exploring social biases inherent in large language models, specifically how these biases affect inclusivity and diversity within AI systems. "I hope to raise awareness of the limitations of AI models and ensure that they are as safe to use as possible," says Hannah.

We are committed to addressing these disparities. Valeria Vitelli, Associate Professor and a Principal Investigator at Integreat, has emphasised the importance of representation across all aspects of life, including cultural backgrounds and gender, to create a more inclusive and equitable research environment.

Our dedication to equity, diversity, inclusion, and belonging extends beyond gender and encompasses all forms of identity and expression. We remain committed to upholding the principles of:

These principles guide our actions and reinforce our ongoing work to create a truly inclusive environment where every individual is valued and empowered. As we move forward, we are excited to continue building a community where everyone, regardless of background, feels they belong to, can thrive, and contribute to the transformative work being done at Integreat. Together, we will continue to create a more inclusive, equitable, and diverse future for research, innovation, and beyond. This is particularly important in the current times.

I - Integrity	Upholding honesty, transparency, and strong moral principles in all actions, ensuring consistency between words and deeds.
N - Non-Discrimination	Treating everyone equally, without bias, ensuring fair and impartial treatment regardless of race, gender, religion, age, identity, or any other characteristic.
T - Tact	Handling sensitive situations with care, understanding the importance of communication, and delivering messages in a thoughtful and respectful manner.
E - Environment	Creating an environment where everyone feels welcome, valued, and respected, ensuring diversity and a sense of belonging.
G - Gratitude	Acknowledging and appreciating the efforts, contributions, and value of others, fostering positive and supportive relationships.
R - Respect	Acknowledging and honouring the rights, values, and dignity of others in all interactions and decisions.
E - Empathy	Understanding and sharing the feelings of others, offering support, and showing compassion.
A - Accountability	Taking responsibility for one's actions, acknowledging mistakes, and learning from them while striving for continuous improvement.
T - Transparency	Being open and clear in communication, fostering trust through honesty, and ensuring that all actions and decisions are visible and understandable.

Seminars 2024



TUESDAY SEMINAR 10.12:

Geir Kjetil
Ferkingstad Sandve

*Using AI to dampen
the unpredictability
brought by climate change
Sustainability series*



TUESDAY SEMINAR 12.11:

Roy Schwartz

*Green AI Sustainability
series*



TUESDAY SEMINAR 22.10:

Volkan Cevher

*Adversarial Training
Should Be Cast as a
Non-Zero-Sum Game*



TUESDAY SEMINAR 15.10:

Aksel Braanen Sterri

*Making AI sustainable
Sustainability series*



TUESDAY SEMINAR 09.10:

Murat A. Erdogdu

*Feature learning in two-layer
neural networks under
structured data*



TUESDAY SEMINAR 17.09:

Vasilis Gkolemis

*Effector: A Python package
for regional explanations*



TUESDAY SEMINAR 03.09:

Tonje Nerby

*AI, scarce resources
and climate transition
Sustainability series*



TUESDAY SEMINAR 27.08:

Claudia Czado

*Vine copula-based
regression models*



TUESDAY SEMINAR 20.08:

Adín Ramírez Rivera

*Learning representations
from high-level structures*



TUESDAY SEMINAR 28.08:

Eric Laber

*Reinforcement learning for
respondent-driven sampling*



TUESDAY SEMINAR 30.05:

Julian Rodemann

*Explaining Bayesian
optimization by
Shapley values*

*Fair, Explainable, and
Trustworthy Learning series*



TUESDAY SEMINAR 16.04:

Mohammad
Emtiyaz Khan

The Bayesian learning rule



TUESDAY SEMINAR 05.03:

Christopher Nemeth

*All models are wrong, and
sometimes, the optimiser
is wrong too!*



TUESDAY SEMINAR 20.02:

Salvatore Florio

*Core Constructional Ontology:
A Unified Theory of Sets,
Parts, and Relations*



TUESDAY SEMINAR 06.02:

Mor Geva

*Knowledge in LLMs:
Lack of Dependencies
and Sparks of Reasoning Fair,
Explainable, and Trustworthy
Learning series*



TUESDAY SEMINAR 23.01:

Rami Mochaourab

*Toward trustworthy AI
- exploring the trade-offs
between accuracy, privacy,
and explainability*



TUESDAY SEMINAR 12.12.23

Ana Ozaki

*Extracting Rules from
Machine Learning Models
in Angluin's Style*

The way ahead

The road in front of us is clear, at least in some sense. From a more general perspective, the horizon appears more blurred.

We expect all projects to deliver the first results, some with groundbreaking potential. Other projects may need to be re-directed. Also, several new projects will start, also in connection with the new recruitments that will happen in 2025: hopefully, we will welcome several Postdoctoral Fellows and new PhD Fellows. In this way, we will be able to support some of the research themes that currently have less activity. We will work further towards close collaborations between our two hubs at UiO and UiT. The Research Themes will start to have a good overview of their activity in Integreat and will be able to start strategic planning. Some will require some modifications because of new breakthroughs in ML.

In the next year, we will investigate the way we are responding to our four objectives.

We expect Integreat to be more visible internationally, through our published results and participation in conferences and by means of alliances with similar centres worldwide.

Our mentoring system will be in place, some research kitchens have been organised.

The Scientific Advisory Committee will visit Integreat again in person in 2025, and we will have our second retreat in the autumn 2025.

In 2025, a series of new AI-research centres will be announced, and Integreat is centrally involved in several. Independently from the results, Integreat will work in a new national landscape, and will need to collaborate with several of the new centres.

We plan to be even more present on the public AI scene in Norway, including the organisation of the Digitalisation-lecture 2025, and an event (or two) at Arendalsuka.

Integreat reflects on the world around us and is directly influenced by events and uncertain perspectives. The wars are not stopping, the alliances are weakening, the economy is less global, there is a renewed need to fund the defense at the cost of other areas, and research and higher education could risk being targeted. In the USA, universities see their federal funding reduced, and the freedom of research is threatened. A fear of challenging security lead many to believe that it is best to isolate research within its own borders, reducing international recruitment and collaboration, which are essential for efficient and successful research. Science is challenged by populist deniers. In most of these aspects, AI plays a very important role. In Integreat, there will be voices against these waves of obscurantism and regression.

Integreat voices

One of Integreat’s key goals for 2024 has been to create an intellectually stimulating and supportive environment where researchers—especially early-career scientists—can thrive both professionally and personally. By prioritising flexibility, work-life balance, community, and mutual respect, Integreat is fostering a work culture that contributes to both productive research and innovative publications.



Emil ↑



Kushagri ↑

A Day in the Life of a PhD Fellow at Integreat

Each day at Integreat is shaped by research, collaboration, and academic engagement. For PhD Fellows like Emil Poiesz and Kushagri Tandon, structuring the day effectively allows for both deep focus and valuable interaction with peers.



Morning: Setting the Tone for the Day

Emil arrives early, enjoying the quiet hours in the beautiful surroundings on the 12th floor at the Integreat premises where half of the PhD Fellows have their desks, before the office fills up with the rest of the people. “I come in around 7 AM when it’s peaceful. It’s my favourite time of the day because I can set my goals and get into the right mindset for work.” He organises his time into large blocks, grouping meetings together to leave uninterrupted space for research and writing.

Kushagri follows a different rhythm, starting her day by reviewing her calendar and emails. “I check my meetings, what I need to prepare, and what tasks I need to complete. These days, there are many meetings, but I always make time for project work, coursework, and reading groups.”



Midday: Research, Writing, and Collaboration

As the day progresses, Emil and Kushagri dive into research. Emil, who has taken on a heavier course load this semester, balances classes with his own project work. “This semester is intense, but it means I’ll have more freedom later. I also assist with teaching, updating course materials and lecture slides—it’s a different but valuable kind of work.”

Beyond structured research work, the day at Integreat is also shaped by spontaneous interactions and collaborative exchanges. There is always an opportunity for informal talks with fellow PhD researchers, whether discussing shared challenges or brainstorming solutions to complex problems. These moments of discussion often spark new insights and perspectives.

A break in the lounge may lead to an impromptu meeting with a supervisor or colleague, where science is explored from different angles. “Some of the best ideas come from these casual conversations, where we see new ways to approach our research questions,” Kushagri reflects.



Afternoon: Engaging with the Research Community

Beyond individual research, Integreat emphasises knowledge exchange through reading groups and seminars. Kushagri is actively involved in several groups, including an online seminar at University College London, an international reading group, and the UiO Bayesian reading group. “Attending seminars and discussions keeps me connected with ongoing research and helps me refine my work.”



Evening: Work-Life Balance and Reflection

Maintaining a work-life balance is essential for long-term success. Emil, who prefers early mornings, wraps up his day by mid-afternoon. “I leave around 3 or 4 PM and shift gears completely. Swing dancing is my way to unwind—it helps me reset and stay energised.”

Kushagri also values balance but notes that adjusting to the Norwegian work culture was an initial challenge. “Work-life balance here is very different from India, but I appreciate how much personal time is respected. I’ve learned to manage my time well while taking necessary breaks.”

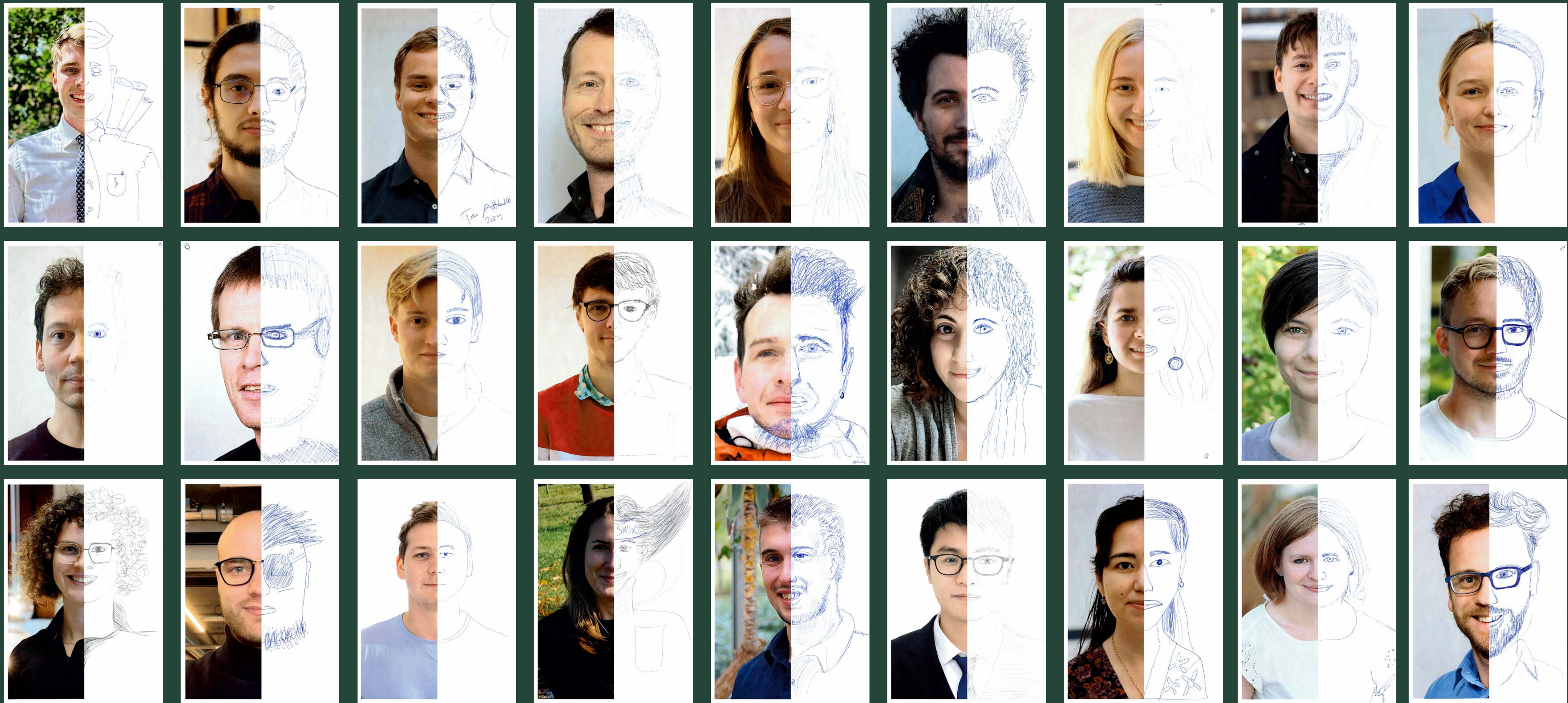


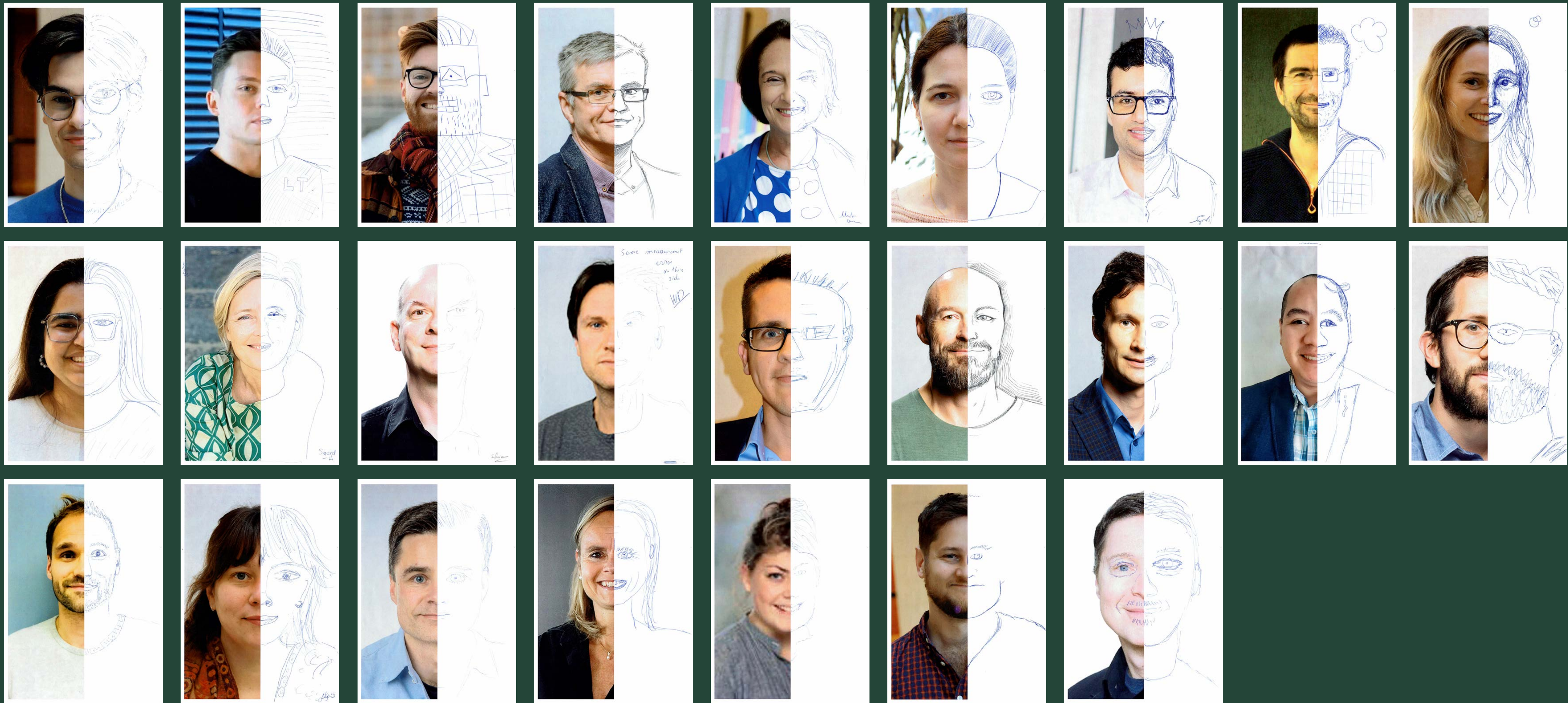
Creating a Thriving Research Environment

Integreat’s commitment to fostering an engaging and supportive academic environment is evident in its approach to structured research activities, collaborative opportunities, and work-life balance. By providing the flexibility and resources needed for both professional and personal growth, the centre continues to create a thriving space where researchers can excel.



The Integreat community





Partners



UNIVERSITY
OF OSLO



UiT The Arctic
University of Norway



NR

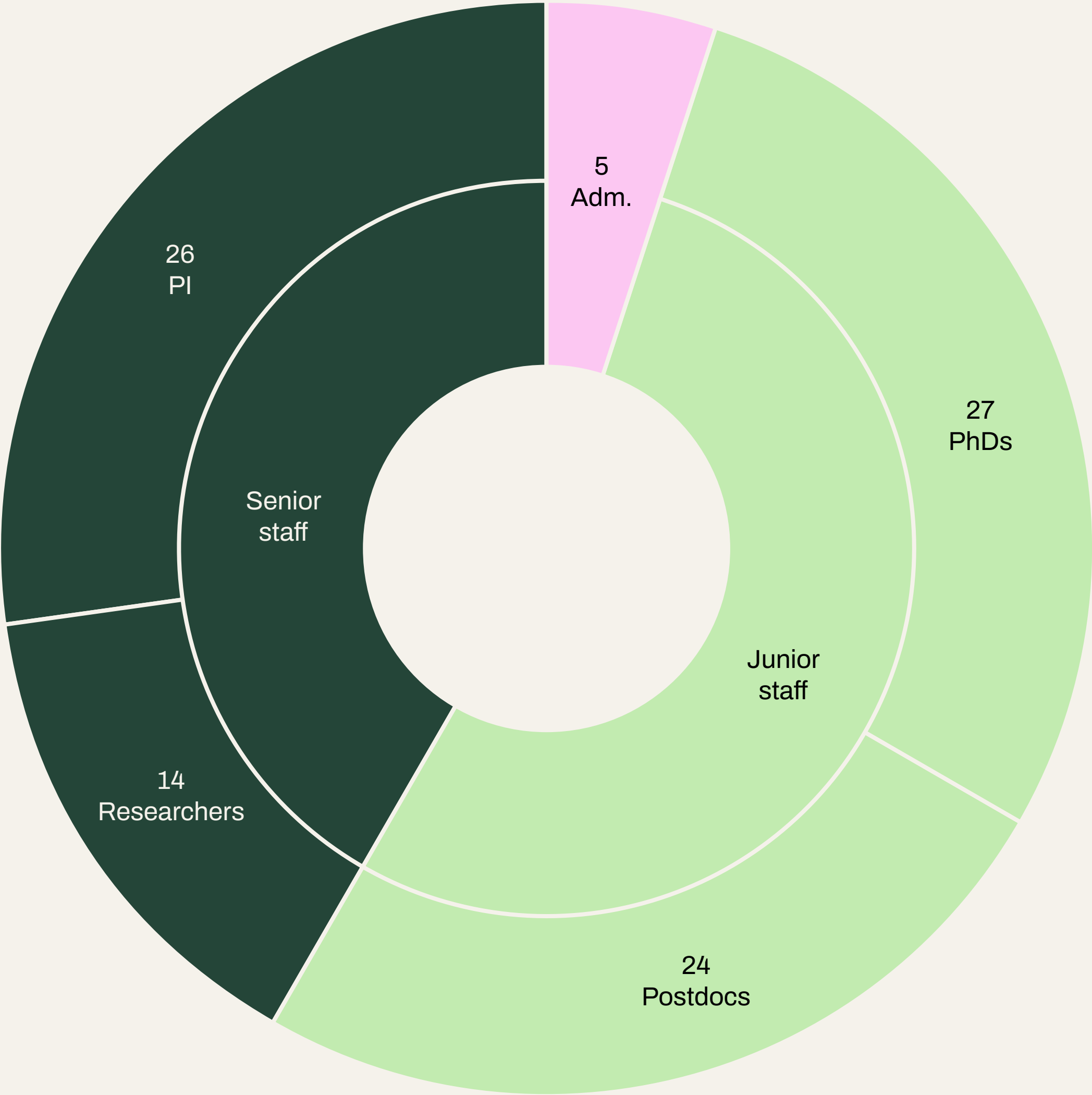
**NORSK
REGNESENTRAL**
NORWEGIAN
COMPUTING CENTER



Norwegian
Centre of
Excellence

People

Staff distribution
by the end of 2024.



Integreat in numbers

September 2023
→ March 2025

} 100
Researchers

} 25
Principal
investigators

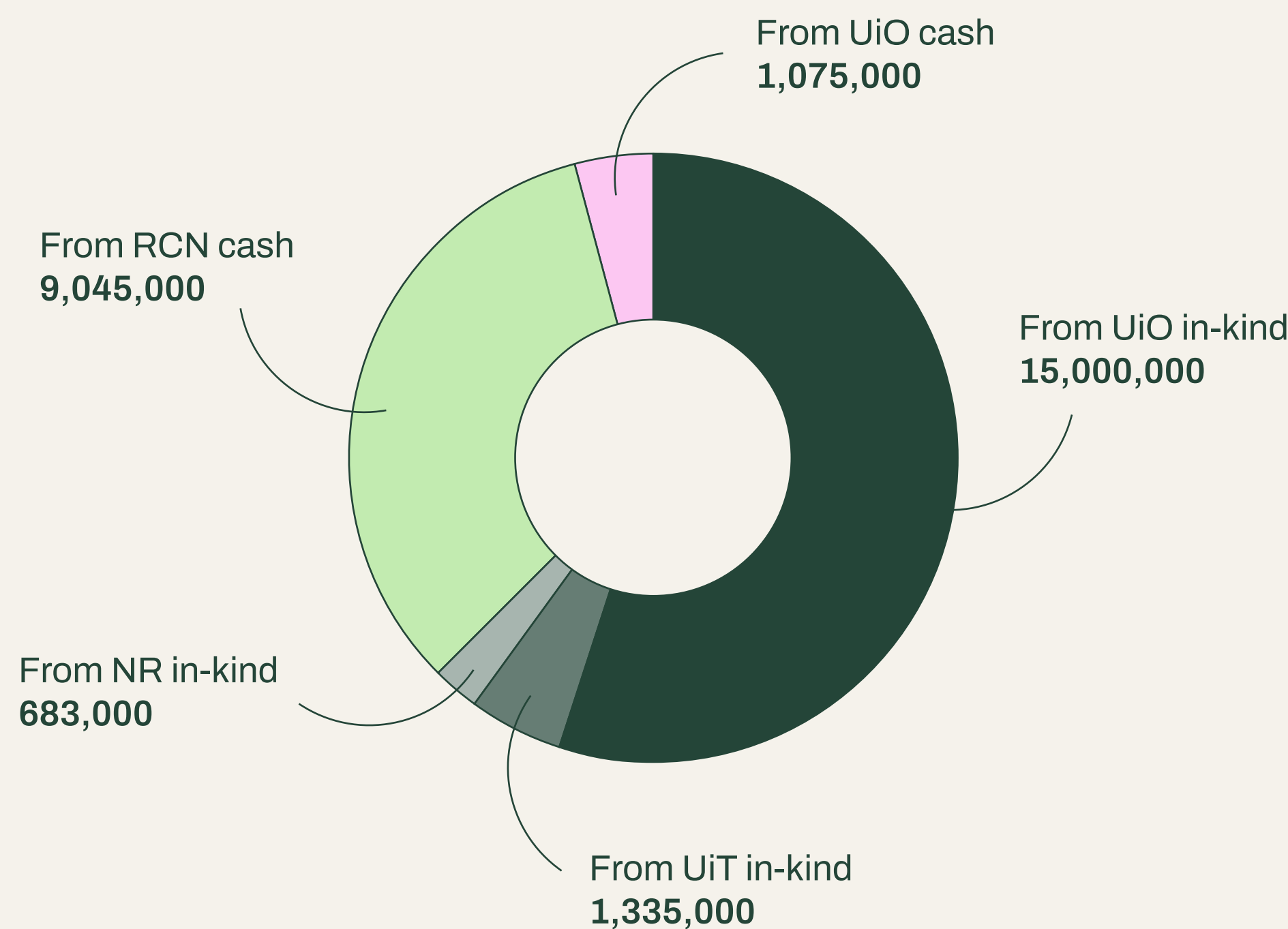
} 30
PhD and
postdocs

} 20
Research
seminars

} 45
Projects

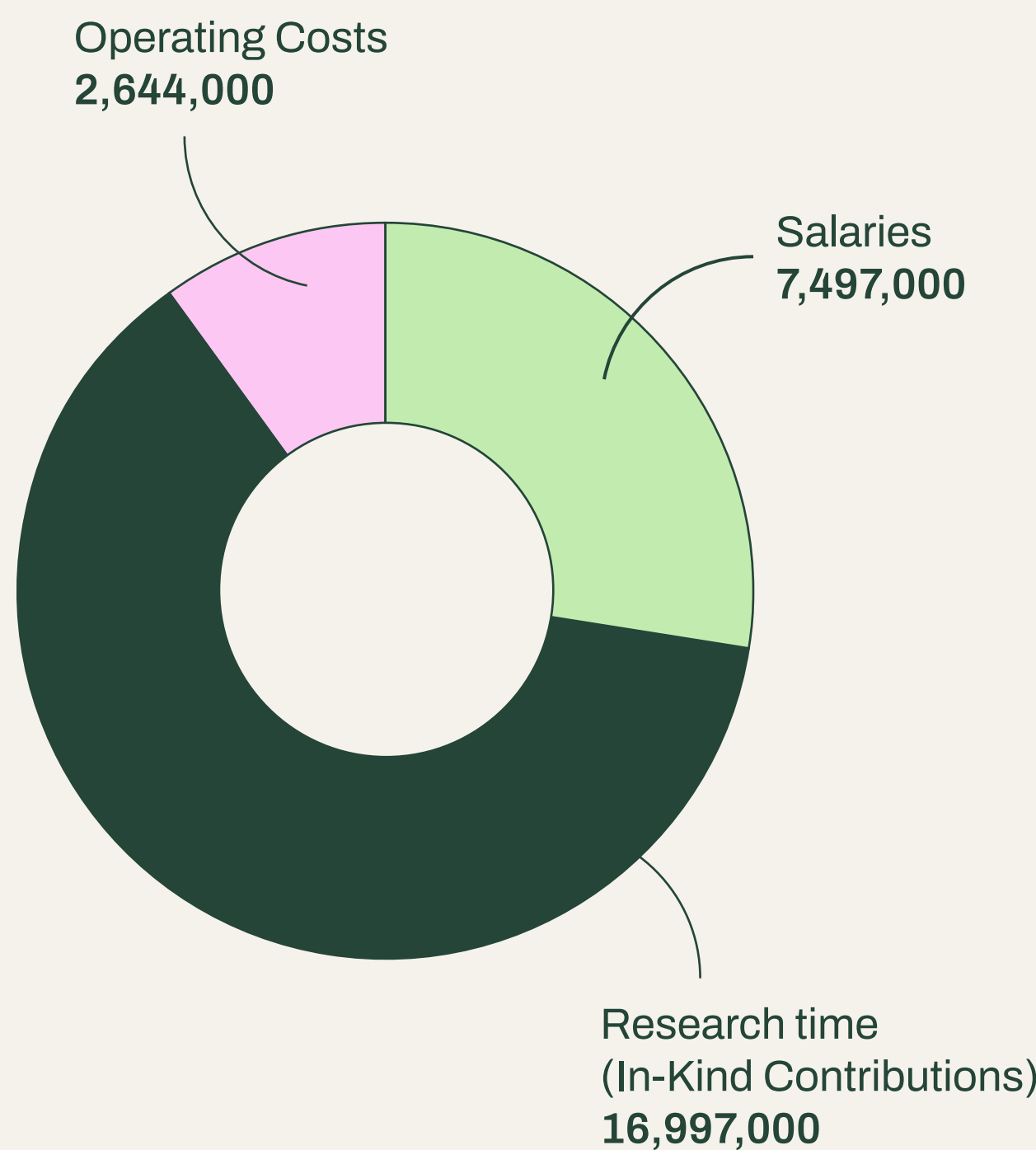
Summary of the fiscal year (2024)

Revenue:



Total revenue cash, total revenue in-kind contributions: 27 138,000

Costs



Total Costs: 27,138,000

Funded by the Research Council of Norway,
Norwegian Centres of Excellence scheme,
Integreat - Norwegian Centre for knowledge-driven machine learning,
project number 332645

Photos: Christoffer Hals, Petter Bjørklund, Olesia Fesenko and Maria Dikova
Design/Layout: Anagram Design AS

Visiting address Oslo

Niels Henrik Abels hus
Moltke Moes vei 35
0851 OSLO
Norway

Visiting address Tromsø

Sykehusvegen 21
Forskningsparken
9019 TROMSØ
Norway

Contact

+47 228 58 898
admin@integreat.no
integreat.no

Follow Integreat on [LinkedIn](#)

integreat