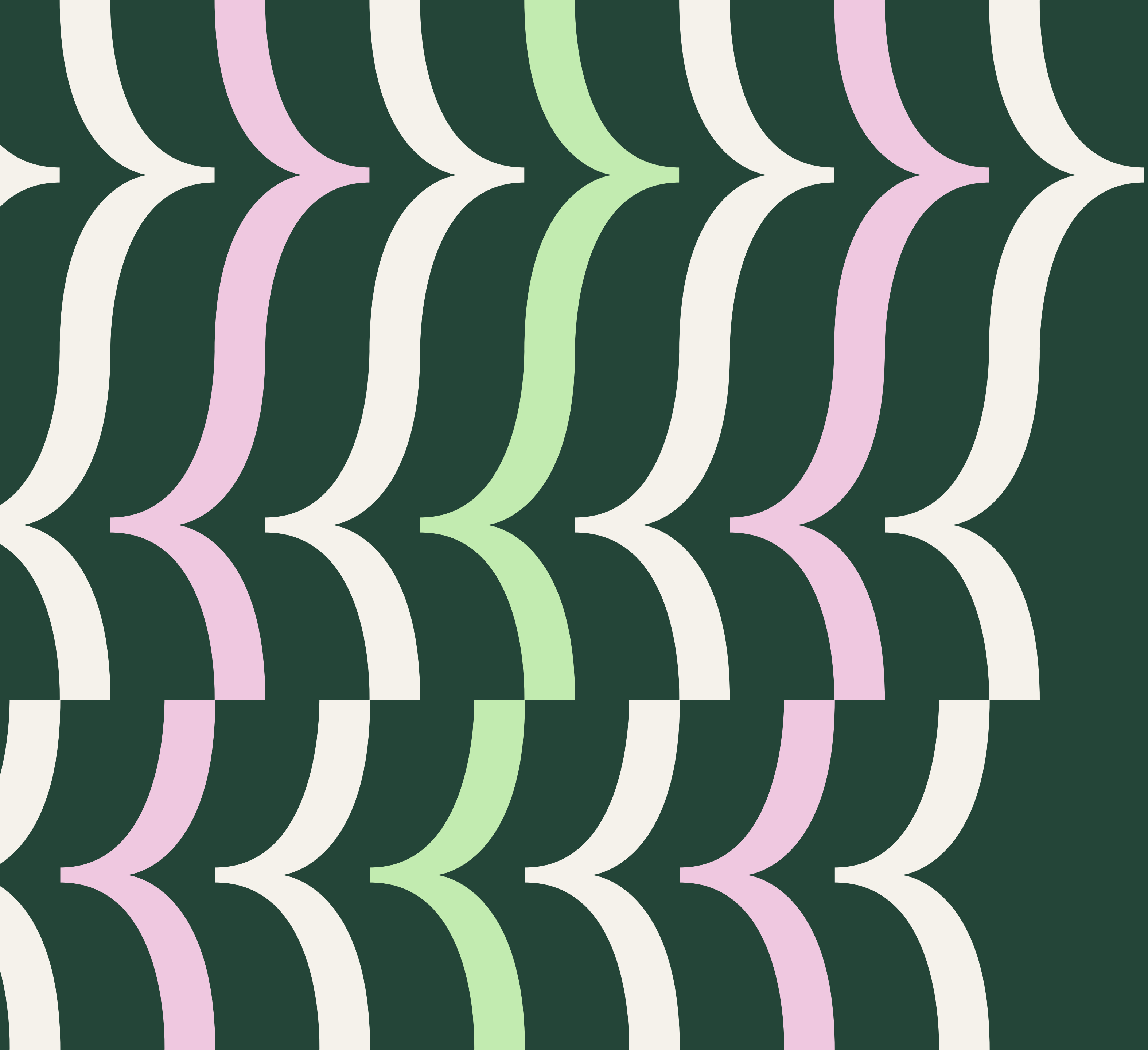




Annual Report 2025

integreat

The Norwegian centre for
knowledge-driven machine learning

Content

About Integreat	5	International collaborations		DEI	71
Organisation	6	and mobility	41	In conversation	
Research Organisation	13	Integreat voices	46	with practice	75
Research highlights	15	Events	50	Photowall	80
Projects	34	Meeting the public	62	Integreat in numbers	82
Publications	38	Researcher training	68	Partners	83

In motion



Integreat Centre Director
Ingrid K. Glad

Oslo, February 2026

Welcome to the 2025 annual report of Integreat, the Norwegian Centre for Knowledge-driven Machine Learning. If 2024 was defined by “building, bonding, sowing, and strategising”, the year of 2025 has been the year of consolidation, when we can see the first tangible results of our efforts and the start of harvesting. Our structures are now well established, and many research projects are in full swing. The intellectual energy across our hubs in Oslo and Tromsø has gradually developed into sustained interdisciplinary collaboration. This has generated a steady stream of results, some of which may shape future directions in the field.

Integreat is an ambitious and friendly home for a growing community of approximately 100 researchers. We host a large, active cohort of younger researchers, who are the lifeblood of our interdisciplinary centre. Their work ensures that our core mission – to make machine learning more precise, sustainable, fair, and with quantified uncertainty – remains at the forefront of our engagements, within the omnipresent global AI conversation.

We continue to merge inductive learning, deductive logic, and inferential statistics. In 2025, we took significant steps to strengthen our focus on causal learning and AI ethics through targeted recruitment and additional resource allocation. Highlights from this period showcase our commitment to solving central open problems. For example, our work on *Fluent Alignment with Disfluent Judges* introduces a fine-tuning method for low-resource languages (such

as Norwegian). It is groundbreaking in its use of self-generated data and “disfluent” judges – who understand but cannot produce the language – enabling instruction tuning without curated datasets. *CogniLoad* is our diagnostic benchmark that breaks new ground by being the first to distinguish between task complexity, distractor interference, and task length to accurately determine why large language models fail at logical tasks. Our new method *TVineSynth* proposes a pioneering truncated C-vine copula generator that better balances privacy and utility when synthesizing tabular data. Our work on *rewards and AI ethics* demonstrates how we define “success” in an algorithm (the reward function) is as critical to user autonomy and safety as the data itself. These are just some examples of the splendid work that is currently done at Integreat: we work systematically and ask ourselves when new results crystallise: **what ground is our result precisely breaking? What is the potential impact of the result, in depth and breadth. Why should the result attract international attention?** To answer these questions is not easy; citations and publication channels are not necessarily relevant indicators, at least not at this early stage.

Our commitment to open AI research remains fundamental. In a world where AI power often concentrates in the hands of a few, Integreat stands for a transparent and democratic development. Our international profile increases steadily: we host prominent international guests, present at the world’s leading conferences, and one of our PIs is now leading a prestigious EU Synergy grant. The recognition of our work is also evident domestically. In 2025, three of our PIs were promoted to Professor. Three of our projects received funding from FRIPRO, the Research Council of Norway’s open and competitive funding scheme for researcher-initiated projects across all disciplines.

The Integreat people and community generate impact that extends far beyond our research results. Our researchers are active participants in the ongoing public conversation about artificial intelligence and

its role in society. At Arendalsuka, a group of Integreat’s early-career researchers presented their work and engaged in discussion with youth politicians, bringing research perspectives directly into public debate. Through initiatives such as our new mathematics course for teachers in Oslo, we also contribute to strengthening understanding of AI in education and practice.

The annual Digitalisation Lecture, organised by Integreat at the University of Oslo, has developed into a recognised arena for dialogue between research and national policy. By inviting the Minister for Digitalisation and Public Governance to take stock of Norway’s digital development, the lecture bridges academic insight and governmental priorities.

Our commitment to “standing up for science” was reflected in the support for the US researchers: together with the Bergen SFF Center for Digital Narrative, we expressed public support for our colleagues in the USA with a *#StandupForScience2025* demonstration to draw attention to the attacks on science happening in the USA. In June, Integreat celebrated Pride Month as an intrinsic part of the centre’s work on equity, diversity, and inclusion. We wish to create an inclusive research environment where everyone can be themselves and contribute to their full potential.

As we look ahead, we are excited by the launch of TRUST, one of the new government-supported AI centers. Built on the foundation of Integreat’s pioneering work and supported by over 40 user partners from diverse perspectives, TRUST represents a new chapter in our journey and exciting opportunities for impactful collaboration and reciprocal benefit.

Since 1 January 2026, I have had the pleasure of directing Integreat following Arnoldo Frigessi, who has moved on to lead TRUST. We would like to thank Arnoldo for his tremendous work in building Integreat, and we are pleased that he will continue with Integreat as a co-director.



*Integreat is building a whole new scientific community across disciplines and scientific traditions. It shows that it is possible to do research in new ways with new colleagues – **remarkable results are happening just now**. Academic environments must dynamically evolve to be able to embrace centres like Integreat, and we need institutions to do more to favour it.*

– Arnaldo Frigessi, Integreat Centre Director 2023–2025

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 bringing together researchers from the University of Oslo, the Norwegian Computing Centre (NR), and Oslo University Hospital, and a hub in Tromsø at UiT – The Arctic University of Norway.

Building on this foundation, Integreat’s research has now moved fully into its project phase. Across themes and disciplines, we are working to integrate general and domain-specific knowledge into machine learning, moving beyond purely data-centric approaches. By combining statistics, language technology, logic, mathematics, ethics, and machine learning, the centre develops theories, models, and algorithms that aim to make ML more accurate, explainable, trustworthy, and sustainable, while improving how uncertainty is understood and quantified.

Interdisciplinary collaboration shapes both the questions we ask and the methods we develop. With strong international collaborations and close engagement across research environments, Integreat is consolidating its role as a centre where complex problems in science, technology, health, and society are approached through knowledge-driven ML.



Organisation

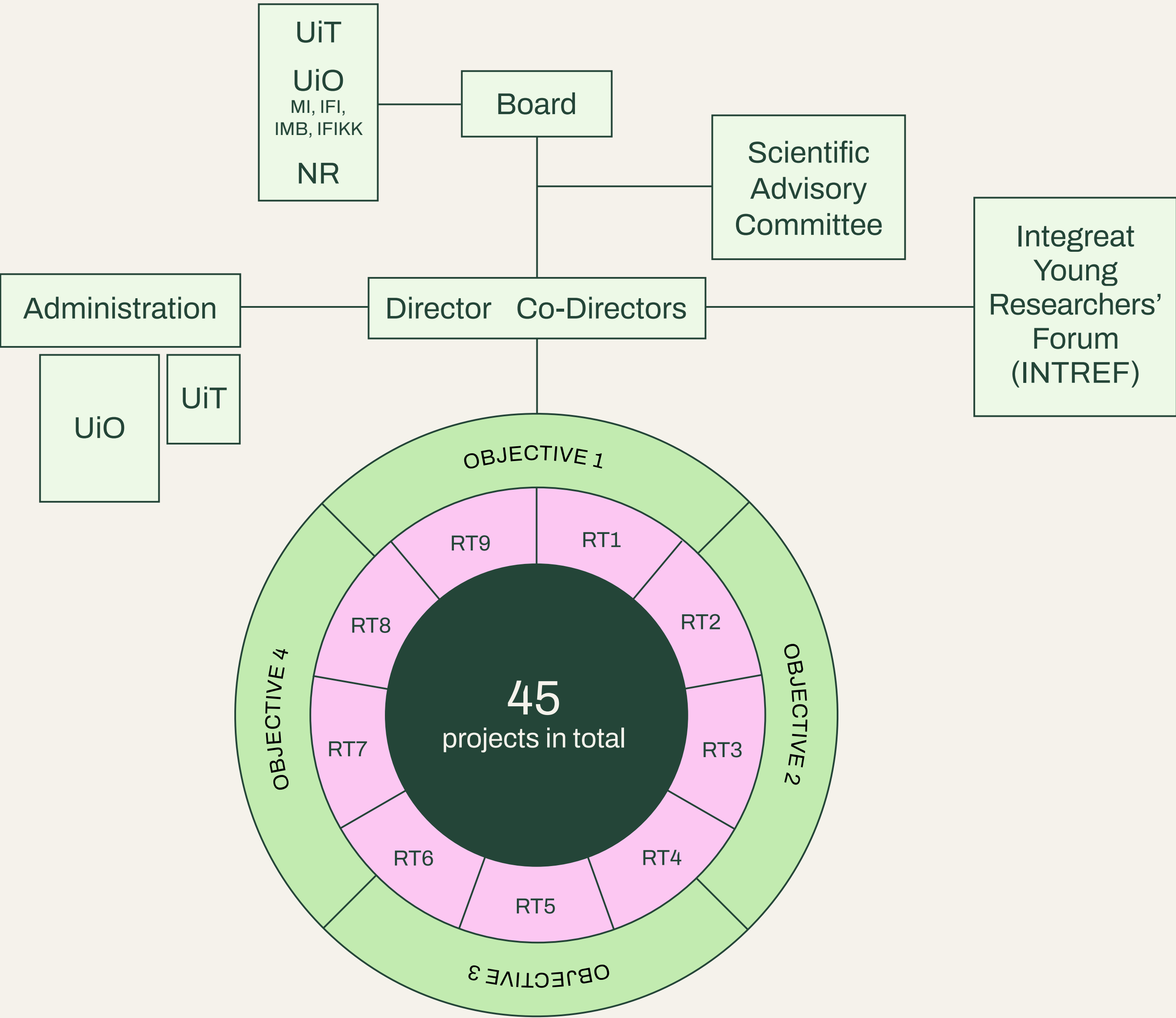
The centre was established in September 2023. As of early 2026, it comprises approximately 100 members, including researchers at all career stages – from master’s students to professors – together with a dedicated group of administrative staff.

At the University of Oslo, researchers are affiliated with the Faculty of Medicine (the Institute of Basic Medical Sciences and the Oslo Centre for Biostatistics and Epidemiology), the Faculty of Mathematics and Natural Sciences (the Department of Mathematics and the Department of Informatics), and the Faculty of Humanities, represented by the 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 and as of 2025, also researchers from the Department of Mathematics and Statistics.

In addition, a group of researchers is affiliated with The Norwegian Computing Centre, which constitutes the third key partner of the centre.

Organisation



Management and administration



↑ Ingrid Glad



↑ Arnaldo Frigessi



↑ Lilja Øvrelid



↑ Benjamin Ricaud



↑ Martin Giese

Integreat is led by a group of co-directors:

- **Professor Ingrid Glad, Director,** Department of Mathematics, University of Oslo
- **Professor Arnaldo Frigessi,** the Institute of Basic Medical Sciences and the Oslo Centre for Biostatistics and Epidemiology, University of Oslo
- **Professor Lilja Øvrelid,** Department of Informatics, University of Oslo

- **Associate Professor Benjamin Ricaud,** Department of Physics and Technology, UiT – The Arctic University of Norway, Tromsø
- **Professor Martin Giese,** Department of Informatics, University of Oslo

The group of co-directors 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 is supported by administrative and technical staff across all partner institutions, assisting with procurement, finance, HR, and communication. In 2025, this work was coordinated by a core administrative team of three, working closely with colleagues at UiT and the other partners.

Much of the centre's coherence relies on this small team, which works across institutions and roles.

Integreat Board 2025

CHAIR

Øystein Prytz

Vice Dean of Research,
the Faculty of Mathematics
and Natural Sciences,
University of Oslo

DEPUTY CHAIR

**Olav Gaute
Hellesø**

Head of Department
of Physics and Technology,
UiT – The Arctic
University of Norway

BOARD MEMBER

**André
Teigland**

CEO, Norwegian
Computing
Centre (NR)

BOARD MEMBER

**Geir
Dahl**

Head of Department
of Mathematics,
University of Oslo

BOARD MEMBER

**Ole Christian
Lingjærde**

Head of Department
of Informatics,
University of Oslo

BOARD MEMBER

**Philippe
Collas**

Head of Department,
Institute of Basic
Medical Sciences,
University of Oslo

Scientific Advisory Committee

The Scientific Advisory Committee (SAC) plays an important role in advising the centre on research strategy and scientific quality. Composed of leading international experts in machine learning, statistics, logic, and language technology, the committee provides independent perspectives on research directions, methodologies, and emerging trends in the field.

In September 2025, the SAC met the codirectors in person in Oslo, for a two-day meeting. The group had focused and in-depth discussions of the centre's ongoing research projects. This allowed for thorough and constructive feedback at the project level, as well as broader reflection on the centre's scientific direction and ways to achieve groundbreaking results. The discussions were marked by openness, and the feedback was timely and encouraging. For Integreat, the interaction with the SAC confirmed the strength of recent recruitments and the quality of the projects now underway. At the same time, the discussions highlighted areas that require

further attention, including the continued sharpening of research questions, clearer articulation of project-specific long-term ambitions, and stronger integration across themes. The feedback was direct and intellectually demanding, yet delivered in a collegial and supportive spirit. Both researchers and leadership experienced the meeting as motivating. It underscored the value of a Scientific Advisory Committee that not only evaluates results, but challenges assumptions and helps refine direction. Such dialogue is essential in a centre that is continuously working towards its ambitions.

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



SAC visiting in Oslo, 1–3 September 2025. From the left Martin Giese, Ian Horrocks, Lilja Øvrelid, Barbara Plank, Ingrid Glad, Arnaldo Frigessi, Sofia Olhede, and Benjamin Ricaud.

« *Integreat combines leading expertise in statistics, logic and machine learning with the goal to enable trustworthy knowledge-driven machine learning. Exactly such interdisciplinary nature is what is needed to reach outcomes that can otherwise not be reached by any of the fields working separately. The Centre is precisely the mechanism needed to make societally impactful outcomes possible.* »

— Professor Barbara Plank, SAC member



[Visit webpage](#)

The INTe great Young REsearchers Forum (INTREF)

The Integreat Young Researchers' Forum (INTREF) brings together the PhD candidates and postdoctoral fellows at the centre and serves as a forum for dialogue, representation, and exchange across research themes. INTREF is led by an elected Fellows Committee that represents the interests of its members and contributes to the ongoing development of the centre's research environment.

In 2025, the committee consisted of five representatives and was supported by the INTREF Coordinator and the Integreat Administrative Leader. Through this structure, early-career researchers are given a clear voice within the centre and a shared space to engage with both scientific and organisational questions. Activity report pages [69–70](#).

Committee in 2025

- **INTREF coordinator:**
Associate Professor Ingrid Hobæk Haff,
Integreat 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, Department of Physics
and Technology, UiT
- **Kushagri Tandon**,
PhD Fellow, Department of Mathematics,
UiO
- **Roberto Neglia**,
PhD Fellow, Department of Mathematics
and Statistics, UiT



Diversity, Equity and Inclusion committee

The Diversity, Equity and Inclusion (DEI) Committee consists of the centre's co-directors, the Integreat Administrative Leader, and Integreat's Inclusion Officer Elisabeth Griesbauer in 2025. The committee provides leadership on equity, diversity, and inclusion, and ensures that these considerations are integrated into the centre's everyday practices and longer-term development.

In 2025, the committee led advocacy and awareness efforts aimed at fostering an inclusive and welcoming research environment across identity groups. This

included the development and follow-up of diversity-related policies, the organisation of seminars and events, participation in Pride month activities, and collaboration with relevant internal and external stakeholders.

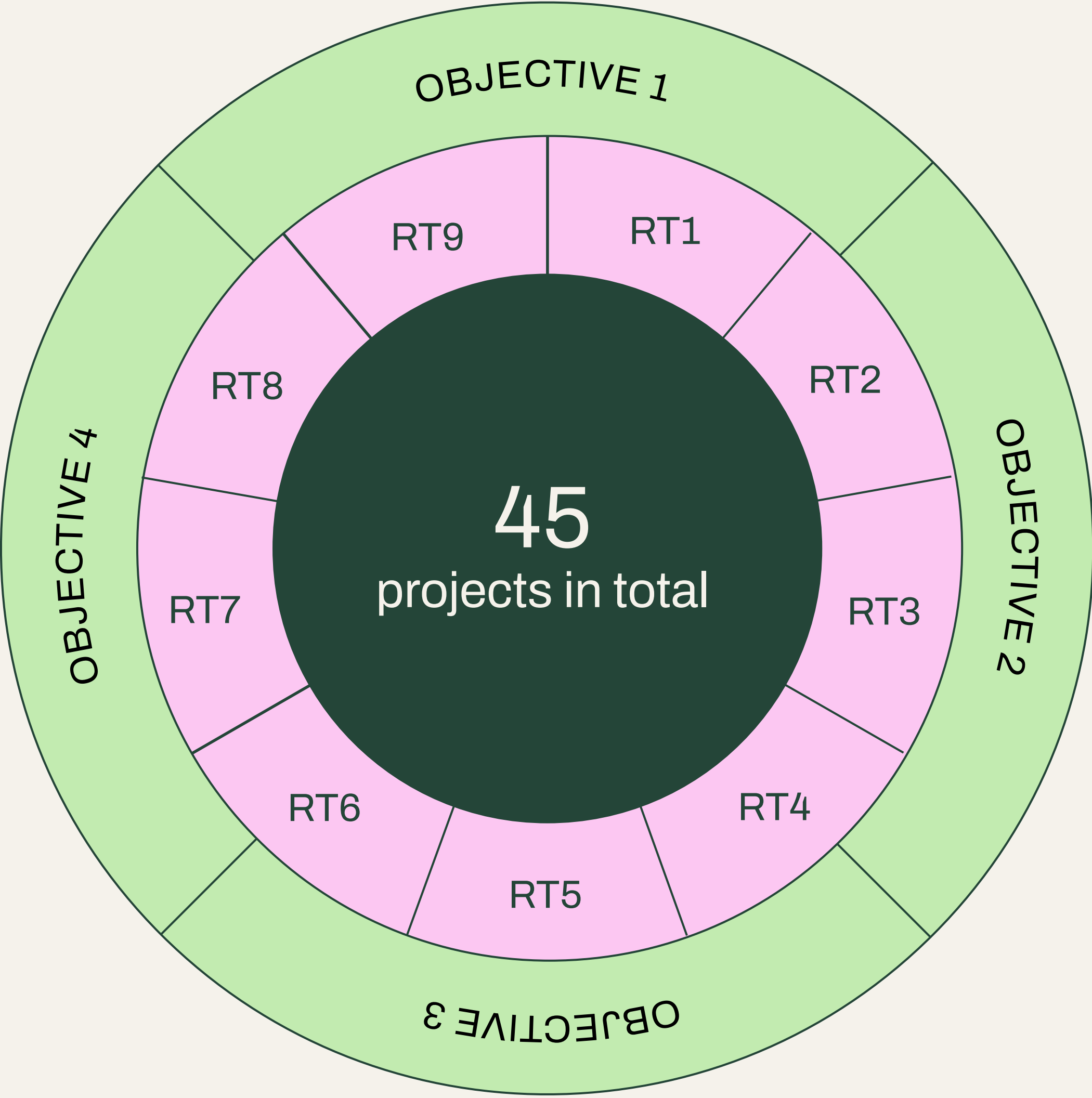
Through this work, the DEI Committee contributes to creating a research community where difference is recognised and supported, and where inclusion is understood as an ongoing responsibility rather than a one-time initiative. [See Pages 71.](#)



Integreat Inclusion Officer 2023-2025 PhD Fellow Elisabeth Griesbauer.

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 2025

Integreat's research is organised into nine interconnected Research Themes (RTs), each addressing specific methods and challenges in knowledge-driven machine learning. Together, the themes span the centre's four overarching objectives and provide a shared structure for interdisciplinary work across domains.

As the centre has moved into its project phase, this organisation has become an active framework. Each RT is guided by a team of Principal Investigators (PIs), who steer scientific direction and oversee research progress, while individual projects increasingly cut across multiple themes. This has required ongoing coordination and dialogue, and has also shaped how research questions are formulated and refined.

Each Research Theme encompasses a diverse portfolio of projects that often intersect across multiple RTs, reflecting our integrated approach.

In this section of the annual report, we highlight the progress made across our RTs during 2025. While a comprehensive list of publications is provided on pp. 38–40, we aim here to give an idea of some of our scientific achievements.

Throughout 2025, our research was featured in premier international venues, including leading conferences such as *NeurIPS*, *ICML*, *ICLR*, *ACL*, *AAAI*, *IJCAI*, *EMNLP*, as well as top-tier journals like *Bayesian Analysis*, *Scandinavian Journal of Statistics*, *Journal of Causal Inference*, *JRSS-A*, *Expert Systems with Applications*.

Beyond publication success, we also celebrate the vibrant academic community surrounding our RTs. We highlight the organisation of events and workshops and welcoming distinguished guest researchers.

We remain particularly energised by our most recent breakthroughs, many of which are currently under submission or available on arXiv – setting a strong foundation for the year ahead.



QR to Research



[Visit webpage](#)

QR to the list of PIs



[Visit webpage](#)

RT1: Probabilistic modelling and uncertainty



[Visit webpage](#)

The mission of this research theme is to investigate the foundational statistical principles that underpin modern AI, and to exploit the inherent stochasticity of most ML algorithms rather than ignoring it. By integrating inferential statistics with probabilistic machine learning, we develop models that provide both rigorous guarantees and the precision necessary for critical, real-world decision-making. Uncertainty quantification and probabilistic modelling remain very central to much research at Integreat, as also visible through the long list of researchers whose interest (also) lies in these areas.

In 2025, RT1 made significant progress in investigating the probabilistic principles that underpin reliable AI, focusing on Bayesian methods and uncertainty quantification, Gaussian processes, robust diagnostics and more. RT1 is closely entangled into all the other research themes; here we mention three projects from RT1 as examples: Leiv Rønneberg’s paper (Rooijakke et al) extends robust and conjugate Gaussian processes

to the multi-output setting, which directly targets uncertainty quantification in systems with several correlated responses by preserving a coherent joint probabilistic model. By constructing a multi-output GP that is both robust to outliers and remains conjugate, it enables more reliable posterior uncertainties that are less distorted by anomalous observations while still being computationally tractable for UQ tasks. Another example is Scheel and Gåsemyr, who developed new graphical and summary diagnostics for Bayesian Hierarchical Models, by assessing “node-level adequacy,” that pinpoints exactly where a complex model may be misaligned with the observed data. Real-world data is rarely static, and our methods must adapt to sudden shifts. Detection of distributional shifts in data streams is the topic of the work by Per August Moen, which proposes a general methodology for fast online changepoint detection. This is essential for AI systems operating in real-time, enabling them to recognize and exploit structural changes in data streams as they happen.

During spring 2025, the Integreat seminar series had a particular focus on uncertainty quantification. It started with Eyke Hüllermeier (LMU) giving a tutorial on distinguishing and representing aleatoric and epistemic uncertainty in ML, followed by Henrik Boström (KTH) on conformal prediction in uncertainty quantification. Then Jeremias Knoblauch (UCL) provided his perspective on Post-Bayesian ML, while Willem Waegeman (Ghent University) discussed making ML models more trustworthy through uncertainty disentanglement. Finally, Henri Prade (IRIT) gave an overview on how possibility theory can be used for interfacing reasoning and learning.

RT1 co-organised the 2nd International Summer School on Preference Learning for Ranking and Ordinal Data (PrefStat), with courses on learning from preference information. PrefStat is a small summer school fully dedicated to a specialised theme, thus allowing to study more deeply the theme, building a more specific network and starting new collaborations.



RT1 TEAM LEADERS

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

RT1 TEAM

- **PhD Fellow Alfonso Diz-Lois Palomares,**
UiO, Department of Mathematics
- **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 Anne-Marie George,**
UiO, Department of Informatics

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

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

- **PhD Fellow Berit Omli Øksnes,**
UiO, Department of Mathematics

- **Researcher Celestino Creatore,**
UiO, OCBE

- **PhD Fellow Christoffer Lingjærde,**
UiO, OCBE

- **Postdoctoral Fellow Clara Bertinelli Salucci,**
UiO, Department of Mathematics

- **Professor Claudia Czado,**
Technical University of Munich

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

- **PhD Fellow Elisabeth Maria Griesbauer,**
UiO, OCBE

- **Associate Professor Emil Stoltenberg,**
UiO, Department of Mathematics

- **Professor Filippo Maria Bianchi,**
UiT, Department of Mathematics and Statistics

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

- **PhD Fellow Hedvig Borgen Reiersrud,**
UiO, Department of Mathematics

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

- **PhD Fellow Håkon Tasken,**
OCBE

- **Professor Ingrid K. Glad,**
UiO, Department of Mathematics

- **Postdoctoral Fellow Johan de Aguas,**
UiO, Department of Mathematics

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

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

- **Research Director Kjersti Aas,**
NR

- **PhD Fellow Kushagri Tandon,**
UiO, Department of Mathematics

- **Research Scientist Lars Henry Berge Olsen,**
NR

- **Postdoctoral Fellow Leiv Rønneberg,**
UiO, Department of Mathematics

- **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

- **Postdoctoral Fellow Per August Jarval Moen,**
UiO, Department of Mathematics

- **Professor Riccardo de Bin,**
UiO, Department of Mathematics

- **PhD Fellow Roberto Neglia,**
UiT, Department of Mathematics and Statistics

- **PhD Fellow Sebastian Grau Nielsen,**
UiO Department of Mathematics

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

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

- **Professor Tero Antero Aittokallio,**
OCBE

- **PhD Fellow Vera Kvisgaard,**
UiO, Department of Mathematics

- **PhD Fellow Vilde Hansteen Ung,**
UiO, OCBE (until autumn 2025)

- **Researcher William Denault,**
Oslo University Hospital, 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 Machine Learning 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.

Among the key challenges in graph learning that has not yet been fully solved is to effectively learn long-range dependencies. Several approaches have been proposed but even evaluating them in a convincing manner is a non-trivial task. Integreat researcher Egor Kostylev et al introduced GLoRa, a synthetic benchmark that allows testing of systems for this ability in a systematic way. They evaluated state-of-the-art systems using GLoRa and concluded that none of them can confidently claim

to learn long-range dependencies well, thus identifying a critical gap.

Another line of research, conducted by Elisabeth Wetzer and Youssef Wally together with other Integreat PIs, focuses on uncovering spatial co-occurrence patterns of objects while addressing the challenge of incorporating semantic relationships between them. For instance, all dishwashers may be categorized more broadly as kitchen appliances and even more broadly as electrical goods, or a Beagle may be classified as a dog, mammal, and animal. Similar hierarchies of concepts arise in biological contexts, such as cell types, where a T helper cell is a kind of T cell, which in turn is a kind of immune cell. Existing spatial analysis tools for biological tissues often fail to integrate such hierarchical knowledge, typically represented as taxonomies or knowledge graphs. This project aims to bridge this gap by developing novel methodologies that leverage logic-based tools in learning tasks. By creating representations of semantic knowledge, the research enables taxonomy-aware association rule mining and enhances downstream machine learning tasks. This fits excellently with recent work by Basil Ell

on graph pattern-based association rules (GPARs) which paves the way for both generative tasks, where a graph is extended, and evaluative tasks, where the plausibility of a graph is assessed. Due to the generality of the concept we believe that it will find applications in many domains.

All research on artificial intelligence is ultimately meant to serve human interests, so human preferences play an important role in many projects at Integreat. In their paper “On Middle Grounds for Preference Statements,” Anne-Marie George and Ana Ozaki investigated a logic-based way of expressing preferences and dealing with divergences in the preferences via a formal notion of a middle ground. Preference statements are of the form *I prefer ‘a’ to ‘b’*, where ‘a’ and ‘b’ are alternatives expressed over some attributes. E.g., in a trolley problem, one can express ‘I prefer to save 1 adult and 1 child to 2 adults (and 0 children).’ They proved theoretical results on the existence and uniqueness of middle grounds. They also provide algorithms for deciding the existence and finding middle grounds.



On the boundary between statistics and logic, RT2 is investigating conflicting, or “inconsistent” information, which is a common phenomenon in real-world data. E.g., humans may state that they prefer apples to oranges, oranges to cherries, and cherries to apples. Both statistical methods and logical formalisms struggle with inconsistent information, and both fields have investigated possible remedies, arriving at superficially very different solutions. In Integreat, Lorenzo Zuccato and his collaborators compare the approaches and their effectiveness and efficiency in order to achieve a common understanding of the trade-offs in play. Results will be presented in the coming year.

› RT2 TEAM LEADERS

- **Associate Professor Ana Ozaki,**
UiO, Department of Informatics
- **Researcher Basil Ell,**
UiO, Department of Informatics
- **Professor Egor Kostylev,**
UiO, Department of Informatics

- **Associate Professor Elisabeth Wetzer,**
UiT, Department of Physics and Technology
- **Professor Martin Giese,**
UiO, Department of Informatics

› RT2 TEAM

- **Associate Professor Anne-Marie George,**
UiO, Department of Informatics
- **Professor Arnaldo Frigessi,**
UiO, Department of Mathematics/OCBE
- **Associate Professor Ida Scheel,**
UiO, Department of Mathematics
- **PhD Fellow Johan Mylius-Kroken,**
UiT, Department of Physics and Technology
- **Associate Professor Johan Pensar,**
UiO, Department of Mathematics
- **PhD Fellow Lorenzo Zuccato,**
UiO, Department of Mathematics
- **PhD Fellow Luca Cibinel,**
UiO, Department of Mathematics

- **Professor Michael Kampffmeyer,**
UiT, Department of Physics and Technology
- **Professor Riccardo de Bin,**
UiO, Department of Mathematics
- **Professor Valeria Vitelli,**
UiO, OCBE
- **PhD Fellow Youssef Wally,**
UiT, Department of Physics and Technology

RT3: Scalable approximation of models and algorithms



[Visit webpage](#)

In RT3, our research focuses on Scalable Approximation – the art of maintaining mathematical integrity while reducing the computational footprint. We move beyond simple “black-box” optimisation to find the optimal balance between accuracy, privacy, and speed, contributing to several of the objectives of Integreat.

Pruning Complexity for High-Dimensional Data is a central theme. The PhD project of August Arnstad investigates sparsification of Bayesian neural networks, with the aim of significantly reducing model size, while retaining performance and uncertainty quantification. Better methodology for large scale Bayesian machine learning will enable more domain knowledge to be implemented in models, and alleviate the computational burden associated with modern deep learning. The development of TVineSynth (Elisabeth Griesbauer et al.) utilises a truncated vine copula to effectively scale generative modeling to high-dimensional tabular data. Truncation allows the model to capture the most critical statistical dependencies while discarding negligible interactions, making it possible to synthesize private, high-quality data at a lower computational cost. As AI

moves into live environments, algorithms must scale across time without exhaustive re-processing. The “fast online changepoint detection” methodology (Per August Moen) provides a general framework for identifying structural shifts in data as they occur. This ensures that models remain adaptive and accurate in dynamic settings without requiring the massive overhead of total re-training. Scaling deep learning often relies on momentum-based optimization. Our results on Generalization of SGD with Momentum (Ali Ramezani-Kebrya) provided the first theoretical conditions for when these approximations generalize successfully, leading to a new algorithm with proven superior generalization, thus providing the “mathematical safety rails” needed to scale the industry’s most popular optimization methods to even larger and more complex tasks.

Sequential Monte Carlo (SMC) – a class of particle-based Bayesian computational methods – is a key computational tool applied to dynamic processes that are only indirectly observed. The simplest versions of SMCs are very easy to implement and can easily work

online when new observations are dynamically collected. SMC algorithms also have several strong theoretical properties. Previously, we applied such techniques to infer time-varying reproduction numbers for infectious disease models under the Covid-19 pandemic. Although applied successfully in that context, it turned clearly challenging to perform inference in high-dimensional latent spaces and to make predictions long ahead in time. A much-used online method for parameter estimation, the *Storvik-filter*, utilizes the existence of sufficient statistics. However, when the time-horizon is large, this method has proven to struggle. Recent offline methods based on Particle MCMC, are computationally slow and do not scale. Our research in RT3, with PhD student Alfonso Diz-Louis Palomares, combines the strengths of the sufficient statistics approach with the Particle MCMC, by updating simultaneously both the latent process and the parameters. Huge improvements in convergence rates are shown empirically for several settings and results will be published in the coming year.



Through these works, Integreat demonstrates that scalability does not have to come at the expense of precision and explainability. Our results help to ensure that AI remains a robust and accessible tool for large-scale, real-world application.

RT3 TEAM LEADERS

- **Professor Geir Olve Storvik**,
UiO, Department of Mathematics
- **Professor Jukka Corander**,
UiO, OCBE

RT3 TEAM

- **PhD Fellow Alfonso Diz-Lois Palomares**,
UiO, Department of Mathematics
- **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
- **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 (until autumn 2025)

- **Associate Professor Ida Scheel**,
UiO, Department of Mathematics
- **PhD Fellow Lorenzo Zuccato**,
UiO, Department of Mathematics
- **Professor Martin Giese**,
UiO, Department of Informatics
- **Postdoctoral Fellow Per August Jarval Moen**,
UiO, Department of Mathematics
- **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
- **Professor Valeria Vitelli**,
UiO, OCBE
- **PhD Fellow Vilde Hansteen Ung**,
UiO, OCBE (until autumn 2025)



*Collaboration across research themes and disciplines is central to our work, enabling us to connect methods, perspectives, and insights in the development of more **robust and trustworthy machine learning**. As the research has entered its project phase, existing connections between research themes have become increasingly visible in practice, while new ones continue to emerge.*

— RT Leader Associate Professor Johan Pensar

RT4: Integrative learning



[Visit webpage](#)

RT4 concerns integrative learning, i.e., how to integrate different forms of knowledge into joint analyses. In a world where data is increasingly multi-modal and fragmented, our mission is to develop a unified framework that seamlessly merges disparate sources of knowledge – from structured logical rules and physical laws to unstructured data streams and human expertise. We believe in the strength of synthesis of different learning paradigms. By interweaving inductive machine learning, which extracts patterns from examples, with deductive reasoning and statistical inference, we create ML approaches that are more than the sum of their parts. This integrative approach allows our models to remain robust even when data is scarce, utilizing prior knowledge to “fill the gaps” while remaining flexible enough to learn from new observations.

These goals are pursued in several Integreat projects, including data-driven tensor factorization models, hybrid models for personalised medicine, and multi-modal

data integration for personalized recommendations. A different challenge of integrative learning emerges when two or more data or knowledge sources carry inconsistent information about the unknown parameters and processes, as we currently study in a project that will give results in the coming year.

One project in RT4 can be used as a paradigmatic example of the work carried out. “Structure-generating models for transition metal” (Luca Cibinel) advances the integrative-learning goal in several ways, all of which are generic and present in many other Integreat projects too:

› Integration of diverse data types. Here, chemical information encoded in different forms and coming from different datasets is combined together: quantum properties and geometrical information about chemical molecules at the metal-complex level (in tabular form) are integrated with quantum properties related

to the atomic composition of the metal complexes (represented as direct and undirected graphs). This happens through a knowledge graph representing the metal complex, provided as an RDF dataset.

› Integration of domain knowledge into machine learning analyses. Using the chemical knowledge representation described above, reconstruct metal complexes using completion probabilities produced by a Bayesian network and will develop generative models, to propose new metal complexes for a specific purpose, in various domains, including drug discovery.

› Integration across disciplines. In this project logic is used to construct the RDF dataset; machine learning and statistical methods are used to address the specific task of metal-complex reconstruction; domain knowledge (in this case, chemistry) is integrated into every stage of the process.



› RT4 TEAM LEADERS

- **Professor Riccardo de Bin,**
UiO, Department of Mathematics
- **Professor Tero Antero 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
- **Associate Professor Anis Yazidi,**
UiO, Department of Informatics

- **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
- **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 Antero Aittokallio,**
Oslo University Hospital,

Institute for Cancer Research
- **Professor Valeria Vitelli,**
UiO, OCBE
- **Researcher William Denault,**
Oslo University Hospital, OCBE

RT5: Transfer learning



[Visit webpage](#)

In 2025, Integreat researchers have focused on transfer learning through the lens of foundation model adaptation, domain shift mitigation, and the geometric structure of latent representations. Our work aligns with the centre’s goal to leverage pre-trained models to save computational costs while ensuring robust performance across diverse downstream tasks.

A major area of inquiry involves adapting large, pre-trained encoders to specific tasks without full retraining. In *Low-Rank Adaptations for increased Generalization in Foundation Model features* (Boe et al), we investigate methods to enhance the generalization capabilities of foundation models. This theoretical work is complemented by practical applications in challenging environments, as seen in our paper *Leveraging Foundation*

Model Adapters to Enable Robust and Semantic Underwater Exploration (Choi et al), where we apply these adaptation strategies to underwater domains. Furthermore, the fundamental mechanics of how these models transfer knowledge are explored in *Systematic Generalization in Language Models Scales with Information Entropy* (Wold et al), which examines the relationship between entropy and generalization capacity.

We addressed the difficulties of transferring knowledge across differing data distributions. The challenge of shifting data statistics is tackled in *Addressing Label Shift in Distributed Learning via Entropy Regularization* (Wu et al), where we propose the Versatile Robust Label Shift (VRLS) method to manage label density ratios

between training and testing environments. Additionally, we address transfer to lower-resource settings in *Fluent Alignment with Disfluent Judges: Post-training for Lower-resource Languages* (David et al), proposing methods to align models without high-quality instruction data in the target language.

Finally, our research explores the underlying structure of latent spaces to improve the transferability of learned features. Our paper *Flatland and Beyond: Mutual Information Across Geometries* (Wally et al) demonstrates that embedding data in hyperbolic spaces can retain more biologically relevant information than traditional Euclidean spaces, thereby enhancing the utility of representations in fields like spatial biology.



› RT5 TEAM LEADERS

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

› RT5 TEAM

- **Professor Adín Ramírez Rivera**,
UiO, Department of Informatics
- **PhD Fellow Adrian Duric**,
UiO, Department of Informatics
- **Associate Professor Aliaksandr Hubin**,
UiO, Department of Mathematics
- **Associate Professor Ana Ozaki**,
UiO, Department of Informatics
- **Associate Professor Andrei Kutuzov**,
UiO, Department of Informatics
- **Associate Professor Anis Yazidi**,
UiO, Department of Informatics
- **Associate Professor Anne-Marie George**,
UiO, Department of Informatics

- **Professor Arnaldo Frigessi**,
UiO, Department of Mathematics/OCBE
- **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
- **Professor Erik Velldal**,
UiO, Department of Informatics
- **Researcher Henri Pesonen**,
Oslo University Hospital, OCBE
- **Associate Professor Ingrid Hobæk Haff**,
UiO, Department of Mathematics
- **PhD Fellow Johan Mylius-Kroken**,
UiT, Department of Physics and Technology
- **Professor Jukka Corander**,
UiO, OCBE
- **Professor Lilja Øvrelid**,
UiO, Department of Informatics

- **Professor Robert Jenssen**,
UiT, Department of Physics and Technology
- **PhD Fellow Sebastian Grau Nielsen**,
UiO Department of Mathematics
- **PhD Fellow Sondre Wold**,
UiO, Department of Informatics
- **Professor Sylvia Richardson**,
University of Cambridge, MRC Biostatistics Unit
- **Professor Tero Antero Aittokallio**,
Oslo University Hospital, Institute for Cancer Research
- **PhD Fellow Tobias Oskar Müller**,
UiO, Department of Informatics
- **Professor Valeria Vitelli**,
UiO, OCBE
- **PhD Fellow Vilde Hansteen Ung**,
UiO, OCBE (until autumn 2025)
- **Researcher William Denault**,
Oslo University Hospital, OCBE
- **PhD Fellow Youssef Wally**,
UiT, Department of Physics and Technology
- **PhD Fellow Zhiyuan Wu**,
UiO, Department of Informatics

RT6: Causal learning

Causal inference aims to identify and quantify causal effects, which unlike statistical associations, describe how a target variable changes in response to an intervention of another variable. Understanding such relationships is central to real-world decision making and scientific discovery, as it concerns the underlying data-generating process and the consequences of interventions (e.g. policy or medical treatment). 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 generalize 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.

Inferring causal relationships from observational data alone is inherently challenging and subject to considerable uncertainty. In this setting, Kvisgaard et al developed a Bayesian procedure for estimating intervention effects between variables in a high-dimensional system. By combining MCMC-based structure learning and local effect estimation, the method was shown to be substantially faster than alternative approaches while achieving comparable or higher accuracy. In a follow-up work, the same researchers showed how prior knowledge about the causal structure

can be incorporated in the learning algorithm to improve the performance. Importantly, the method provides a data-driven tool for identifying plausible causal relationships, e.g. to guide follow-up experiments, and it is currently being extended to also integrate information from interventional data arising from perturbations of the underlying data-generating system.

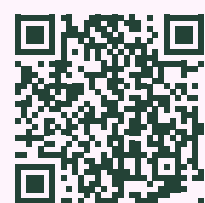
Missing data present a key challenge. In this context, recoverability refers to whether a target quantity can be consistently estimated from incomplete data. The framework of graphical models has recently gained attention as a tool for handling complex missingness mechanisms. Contributing to this line of work, de Aguas et al. extended a popular graph-based missingness model to the important settings where missingness can influence downstream actions, leading to mechanism shifts. They derived criteria for recovering causal effects under the new model and proposed doubly robust estimators. Importantly, this work extends the class of missing-data scenarios that admit recoverability and accurate estimation of causal effects, even when data are not missing at random. More broadly, this work also provides a bridge between causal and agent-based models by introducing causal graphs that encode different agents' world models under varying levels of information access, leading to different perceived data-generating processes and decision rules.

› RT6 LEADERS

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

› RT6 TEAM

- **Postdoctoral Fellow Camilla Lingjærde**,
UiO, Department of Mathematics
- **Professor Geir Kjetil Sandve**,
UiO, Department of Informatics
- **Postdoctoral Fellow Johan de Aguas**,
UiO, Department of Mathematics
- **PhD Fellow Luca Cibinel**,
UiO, Department of Mathematics
- **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



[Visit webpage](#)

RT7: Natural language



[Visit webpage](#)

What we did not predict when we wrote the first research plan of Integreat is the explosion of LLMs, which has 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. In response to these developments, RT7 has been strategically realigned to capture these emerging research directions and the field’s shifting dynamics.

A particular successful research project in 2025 has centered on preference learning for LLMs. This is what governs the behaviour of LLMs – their *style* and *‘personality’* if you will – and is an essential part of what is known as *post-training*, the second of the two main stages of training LLMs. In the first stage, *pre-training*, LLMs acquire their knowledge of the world and their mastery of language by being fed enormous amounts

of raw text. But this alone does not make them useful AI assistants. To work as dialog partners, they also need to understand what makes a response helpful, safe, and appropriate. This is the point of preference-learning. Typically, this is done by further fine-tuning the model on pairs of input prompts and possible responses that humans have annotated as preferred or non-preferred. The process is referred to as *Reinforcement Learning with Human Feedback*, or RLHF. However, this reliance on human annotation creates a data bottleneck, particularly for smaller languages like Norwegian.

Every Wednesday, a bunch of Integreaters sit down together to discuss how to improve LLM preference learning, particularly in data-constrained settings. The participants, comprising PhDs, postdocs, and professors, have backgrounds from several different fields – statistics, informatics, language technology, and more. The mix of perspectives has proved not just fun but very fruitful! We are currently working towards scaling up experiments with so-called RLAIIF – *Reinforcement Learning from AI Feedback* – where human annotations are replaced by large ensembles of LLMs-as-judges, while simultaneously replacing the mathematical framework traditionally assumed for preference learning (the so-called Bradley–Terry

model) with a novel approach we believe will be more powerful (based on the Mallows rank model).

In a closely related effort, documented in a forthcoming ICLR paper (Samuel et al. 2026), we show that replacing the usual reward model assumed in RL with an LLM-as-a-judge (again a form of RLAIIF) can create fluent LLMs for lower-resource languages without any target-language annotations at all!

Moreover, to *evaluate* how well different LLMs align with human preferences, we have also created benchmark data (Mikhailov et al. 2025). Here, human annotators have indicated which of two given LLM responses are preferred with respect to different criteria like safety, truthfulness, and overall quality, along with an explanation of the particular preference. This data can, in turn, be used to test both how well models *simulate* human preferences and how well they can *explain* their own preference decisions.

In 2025, Integreat developed CogniLoad (Kaiser et al), a diagnostic benchmark designed to pinpoint exactly where and why models fail when performing logical tasks. Unlike traditional benchmarks, CogniLoad introduces a nuanced methodology inspired by cognitive load theory. The benchmark breaks new ground by

isolating three distinct variables that impact model performance: (i) Task Complexity: The intrinsic difficulty of the logical operation. (ii) Distractor Interference: The model’s ability to filter out irrelevant information. (iii) Task Length: The impact of long-range context on reasoning accuracy. By decoupling these factors, CogniLoad allows researchers to determine if a model’s failure is due to a lack of reasoning “intelligence” or simply an inability to handle “noisy” environments or long sequences. This diagnostic depth is essential for developing more robust, reliable, and transparent AI systems.

RT7 TEAM LEADERS

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

RT7 TEAM

- **Associate Professor Ali Ramezani-Kebrya,**
UiO, Department of Informatics
- **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 Daniel Kaiser,**
UiT, Department of Physics and Technology
- **PhD Fellow David Samuel,**
UiO, Department of Informatics
- **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
- **Postdoctoral Fellow Meihan Tong,**
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
- **Professor Valeria Vitelli,**
UiO, OCBE
- **Postdoctoral Fellow Vladislav Mikhailov,**
UiO, Department of Informatics
- **Chief Research Scientist Pierre Lison,**
NR

RT8: Explainability and fairness



[Visit webpage](#)

During 2025 the research theme on explainability and fairness has been active across research, dissemination, and community-building efforts.

We are very happy about our well attended academic reading group on the explainability of large language models (LLMs). Explaining why a model produces a certain answer, or what it has learned internally, is important for building trust, ensuring fairness, and supporting responsible use of these systems. Some examples of discussed themes were (i) which tokens of a prompt that are most important for producing an answer and (ii) which training documents were actually used when answering a question.

During 2025, PhD student Johan Mylius-Kroken received the best student paper award at the biannual conference of the Norwegian Association of Machine Learning and Image Analysis for work on neural network

explainability. Many other RT8 publications appeared in impactful venues. On the topic of explainability, Olsen et al. propose improvements to the weighting strategy in KernelSHAP, strengthening the reliability of feature-attribution methods, while Kapar et al. use explainable AI techniques to systematically evaluate the quality and limitations of synthetic tabular data generated by modern models, both presented at XAI. Several RT8 publications directly addressed fairness concerns, including Ozaki et al., who analysed gender bias in BERT-based language models using interpretable decision trees, and Smajdor et al., who explored autonomy, trust, and fairness in medical AI decision-making. In addition, a paper on software for Shapley values is currently under revision, and work is ongoing on asymmetric Shapley values for gene importance assessment in collaboration with researchers at Vrije Universiteit Amsterdam.

} **RT8 TEAM LEADERS**

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

} **RT8 TEAM**

- **Associate Professor Ana Ozaki**, UiO, Department of Informatics
- **Professor Anna Smajdor**, UiO, Department of Philosophy, Classics, History of Art and Ideas
- **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
 - **Research Scientist, Ingrid Dæhlen,**
NR
 - **Professor Ingrid K. Glad,**
UiO, Department of Mathematics
- **Associate Professor Ingrid Hobæk Haff,**
UiO, Department of Mathematics
 - **Associate Professor Ingrid Chieh Yu,**
UiO, Department of Informatics
 - **PhD Fellow Johan Mylius-Kroken,**
UiT, Department of Physics and Technology
 - **Associate Professor Johan Pensar,**
UiO, Department of Mathematics
 - **Research Scientist Johannes Voll Kolstø,**
NR
 - **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](#)

RT9 focuses on examining the ethical and philosophical dimensions of AI, machine learning, and data use both within Integreat, and more broadly, fostering dialogue between technical researchers and ethicists. Over the past year, the team has both expanded and deepened this work through new hires, collaborations, and events that connect philosophy with data science and machine learning.

Over the past year, RT9 has embarked on a number of activities:

Postdoc Yael Friedman started to work on aspects of ethical and philosophical significance in relation to the research being undertaken in Integreat more generally. Yael has participated in a number of activities and initiated several new projects and research ideas. She is actively engaged in identifying further possible project ideas with the other RTs. Anna Smajdor and Yael are working on a project about synthetic data, both in medical and other settings, with links to Anna’s

forthcoming ERC Synergy project on synthetic DNA. We will explore this further later in 2026 with a ‘research kitchen’ involving interested people from UiO, UiT and elsewhere, both machine learners and other researchers. We have submitted an abstract to present some of the work on synthetic data at the Lund Medical Ethics Conference in Sweden, May 2026. Hannah Clausen has continued with her doctoral project on ethical benchmarking of LLMs. She has written her first paper together with her supervisors. The work on the ethical significance of rewards in AI has been continued, and in late 2025, Anna Smajdor, Manuela Zucknick and Arnaldo Frigessi had their paper on ethical aspects of rewards in medical AI accepted. Since then, Anna Smajdor has been working towards an idea for addressing the crossover between rewards in AI and dopamine/serotonin rewards in biological organisms, and a potential link between this and her new ERC Synergy project, starting in Sep 2026. To strengthen our AI ethics team, we welcomed Pooya Zakeri to

Integreat. Erik Velldal, Anne-Marie George and Anna Smajdor submitted an abstract for a special issue on animal ethics in relation to AI. Although our submission was not selected, the editors gave positive feedback and encouraged us to work on this collaboration further. In late 2025 Yael organised a one-day symposium for machine learners and philosophers with an interest in ethics and AI to share work in progress. The event was particularly aimed at graduate students and focussed on offering an informal opportunity to discuss shared interests. The event was very well attended and followed up by the creation of a reading group, which will meet on a regular basis. The first reading group meeting was held in January 2026, with a good number of attendees. Publications for RT9 remain relatively few, but there are several very promising activities in progress, and we anticipate that 2026 will see several more papers published in our theme.



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 Pooya Zakeri,**
UiO, Department of Informatics
- **Associate Professor Samia Touileb,**
UiB Department of Information Science
and Media Studies
- **Postdoctoral Fellow Yael Friedman,**
UiO, Department of Mathematics

Project Portfolio

Advancing Statistical Methods for Global Fits in High-Energy Physics — PhD Fellow Hedvig Borgen Reiersrud; Riccardo De Bin — *Research themes: RT1, RT3*

Bayesian Adaptive Networks via Kernels and Nonparametrics — Postdoctoral Fellow Leiv Rønneberg; Geir Olve Storvik — *Research themes: RT1*

Benchmarking Ethical Reasoning by Large Language Models — PhD Fellow Hannah Clausen; Erik Velldal, Andrey Kutuzov, Anna Smajdor — *Research themes: RT7, RT9*

BRIDGE – Bayesian Rank-Based Unsupervised Integration of Multi-Source Data in Cancer Genomics and the Digital Economy — Postdoctoral Fellow Jan Greve, PhD Fellow Jonas Nordstrøm; Valeria Vitelli, Anne-Marie George, Øystein Sørensen — *Research themes: RT1, RT2, RT3, RT4, RT8*

Causal and Causality-Inspired Machine Learning — Postdoctoral Fellow Johan Manuel de Aguas Pérez; Johan Pensar — *Research themes: RT6*

Causal Discovery Using Directed Graphical Models — PhD Fellow Vera Kvisgaard; Johan Pensar — *Research themes: RT6*

Causal Multi-Task Machine Learning for Enhanced Versatility, Equity and Robustness — Postdoctoral Fellow Camilla Lingjærde; Johan Pensar, Geir Kjetil Ferkingstad Sandve, Sylvia Richardson, Arnaldo Frigessi — *Research themes: RT6, RT8*

Compositional reasoning over graphs with LMs – PhD Fellow Sondre Wold; Egor Kostylev, Erik Velldal, Lilja Øvrelid — *Research themes: RT2, RT7*

Data-Driven Tensor Factorization Models — William Denault, Valeria Vitelli, Tero Antero Aittokallio, Sylvia Richardson, Aliaksandr Hubin — *Research themes: RT1, RT2, RT3*

Developing Novel Information Theoretic Discrepancy Measures — PhD Fellow Li Tianchao; Robert Jenssen, Shujian Yu — *Research themes: RT2, RT5*

Discovering and Generating Complex Dependencies in High Dimensional Data — PhD Fellow Berit Omli Øksnes; Ingrid Hobæk Haff, Thordis Linda Thorarinsdottir — *Research themes: RT1, RT2, RT8*

Discrepancy Measures and Accelerated Likelihood-Free Inference for Simulator-Based Models — PhD Fellow Vilde Hansteen Ung; Jukka Corander, Henri Pesonen, Robert Jenssen — *Research themes: RT3*

Embedded Sufficient Statistics — PhD Fellow Sigurd Holmsen; Ali Ramezani-Kebrya, Arnaldo Frigessi, Sylvia Richardson — *Research themes: RT1, RT3, RT5*

Energy-Efficient and Scalable Deep Learning — PhD Fellow Adrian Duric; Postdoctoral Fellow Christina Runkel; PhD Fellow Tobias Muller; Ali Ramezani-Kebrya — *Research themes: RT3*

Epistemological Tensions in AI and their Effect on Ethics — Postdoctoral Fellow Yael Friedman; Anna Smajdor — *Research themes: RT9*

Exploration and Control of the Inner Representation in Generative AI Models — PhD Fellow Daniel Kaiser; Benjamin Ricaud, Ali Ramezani-Kebrya, Arnaldo Frigessi — *Research themes: RT4, RT5, RT7, RT8*

Flexible Modelling of Extreme Natural Processes — Postdoctoral Fellow Touqeer Ahmad; Thordis Linda Thorarinsdottir, Lena Merete Tallaksen, Kolbjørn Engeland — *Research themes: RT1*

Finding a Middle Ground Based on Lexicographic Preferences — Ana Ozaki, Anne-Marie George — *Research themes: RT2, RT9*

Handling Inconsistencies: A Joint Statistical Learning and Logical Reasoning Framework — PhD Fellow Lorenzo Zuccato; Ida Scheel, Martin Giese, Valeria Vitelli, Arnaldo Frigessi, Anne-Marie George — *Research themes: RT1, RT2*

Inference with the Likelihood Ratio When Parameters Are on the Boundary — Postdoctoral Fellow Clara Bertinelli Salucci; Riccardo De Bin — *Research themes: RT1, RT3*

Knowledge and Bias Extraction from Large Language Models — PhD Fellow Emil Poiesz; Ana Ozaki, Lilja Øvrelid, Samia Touileb — *Research themes: RT2, RT7, RT8*

Knowledge-Driven Representation and Similarity Learning — PhD Fellow Youssef Wally; Elisabeth Wetzler, Benjamin Ricaud, Martin Giese, Vladan Milosevic — *Research themes: RT2, RT4, RT5*

Knowledge-Informed Spatio-Temporal Models — PhD Fellow Kushagri Tandon; Thordis Linda Thorarinsdottir, Geir Olve Storvik — *Research themes: RT1, RT2*

Learning Differential Models (and Other Mathematical Structures) from Data with Uncertainty — PhD Fellow Håkon Tasken; Arnaldo Frigessi, Benjamin Ricaud, Alvaro Köhn-Luque, Martin Giese — *Research themes: RT1, RT2, RT6, RT7*

Learning from Human Preferences in Large Language Models — PhD Fellow David Samuel; Erik Velldal, Lilja Øvrelid, Anne-Marie George, Valeria Vitelli, Vladislav Mikhailov, Andrey Kutuzov, Ida Scheel, Arnaldo Frigessi — *Research themes: RT1, RT7, RT9*

Learning Stable Matchings — Anne-Marie George, Christos Dimitrakakis, Andreas Athanasopoulos — *Research themes: RT2, RT4*

Machine Learning in Real World — Postdoctoral Fellow Christina Runkel, PhD Fellow Tobias Mueller; Ali Ramezani-Kebrya — *Research themes: RT3*

Machine Learning for Graphs and Time Series Data — PhD Fellow Roberto Neglia; Filippo Maria Bianchi; Carlo Abate; Andrea Cini — *Research themes: RT1, RT3, RT4*



Multi-Modal Data Integration for Personalized Treatment Recommendations — PhD Fellow Christoffer Lingjærde; Valeria Vitelli, Basil Ell, Tero Antero Aittokallio, Riccardo De Bin, Sylvia Richardson — *Research themes: RT1, RT2, RT4*

On Sequential Monte Carlo Methods in High-Dimensional Settings — PhD Fellow Alfonso Diz-Lois Palomares; Geir Olve Storvik, Ida Scheel — *Research themes: RT1, RT3*

On the Intersection between Information Theory and Deep Learning — PhD Fellow Johan Mylius-Kroken; Kristoffer Wickstrøm, Ali Ramezani-Kebrya, Robert Jenssen, Elisabeth Wetzer — *Research themes: RT1, RT3, RT8*

Probabilistic Preference Learning with the Mallows Model — Arnaldo Frigessi, Valeria Vitelli, Ida Scheel, Øystein Sørensen — *Research themes: RT1, RT3*

Robust Online Anomaly Detection in Non-Stationary Data Streams — PhD Fellow Sebastian Grau Nielsen; Ingrid Kristine Glad, Michael Kampffmeyer, Martin Tveten, Idris Eckley, Per August Jarval Moen — *Research themes: RT1, RT3, RT5*

Separation of Linguistic and Factual Knowledge in Large Language Models — Postdoctoral Fellow Meihan Tong; Andrey Kutuzov, Egor Kostylev — *Research themes: RT7*

Statistically Optimal and Computationally Efficient Methods in Changepoint Analysis — PhD Fellow Per August Jarval Moen; Ingrid Kristine Glad, Martin Tveten — *Research Themes: RT1, RT3*

SODA: Statistical Methods for Online Detection of Anomalies — Postdoctoral Fellow Per August Jarval Moen; Ingrid Kristine Glad, PhD Fellow Sebastian Grau Nielsen, Nils Lid Hjort, Martin Tveten, Idris Eckley — *Research themes: RT1, RT3*

Sparse models for machine learning from a Bayesian viewpoint — PhD Fellow August Arnstad; Geir Olve Storvik, Fred Godtliebsen, Leiv Rønneberg — *Research themes: RT1, RT2, RT3*

Stepwise variational inference with vine copulas — PhD Fellow Elisabeth Griesbauer; Ingrid Hobæk Haff, Arnaldo Frigessi, Claudia Czado, Leiv Rønneberg — *Research themes: RT1, RT2, RT3*

Structure-Generating Models for Transition Metal Complexes — PhD Fellow Luca Cibinel; Basil Ell, Johan Pensar, Riccardo De Bin — *Research themes: RT1, RT2, RT3*

Transforming real-world evidence with large language models: predictive analytics for improved healthcare outcomes — PhD Fellow Michele Chini; Arnaldo Frigessi, Simen Eide, Christian Jonasson, Kristian Svendsen — *Research themes: RT1, RT2, RT3*

Uncertainty quantification in the presence of logical constraints — PhD Fellow Lorenzo Mastrandrea; Egor Kostylev; Arnaldo Frigessi — *Research themes: RT1, RT2*

What's in the Black-Box? Improving the Shapley Value Explanation Methodology — PhD Fellow Lars Henry Olsen; Ingrid K. Glad, Anders Løland, Kjersti Aas — *Research themes: RT1, RT8*



PI Ali Ramezani-Kebrya with Master students

Research in focus

Understanding the Limits of Large Language Models



↑
PhD Fellow Daniel Kaiser

In 2025, PhD Fellow Daniel Kaiser, based at Integreat’s UiT hub, and affiliated with the Machine Learning Group at UiT The Arctic University of Norway, together with the team of supervisors, developed a new method for analysing how large language models (LLMs) solve logical problems. The work was accepted for publication at the International Conference on Learning Representations (ICLR) 2026.

The method, named *CogniLoad*, is designed to examine how models reason when confronted with new information. Rather than testing factual knowledge, *CogniLoad* evaluates how well LLMs track and update information across a sequence of logical statements. The benchmark consists of structured logical riddles whose length, intrinsic difficulty, and level of distracting information can be systematically adjusted.

The design is inspired by Cognitive Load Theory from human cognition research. By varying the amount and

type of information a model must process, the method makes it possible to analyse how different forms of “cognitive load” affect reasoning performance.

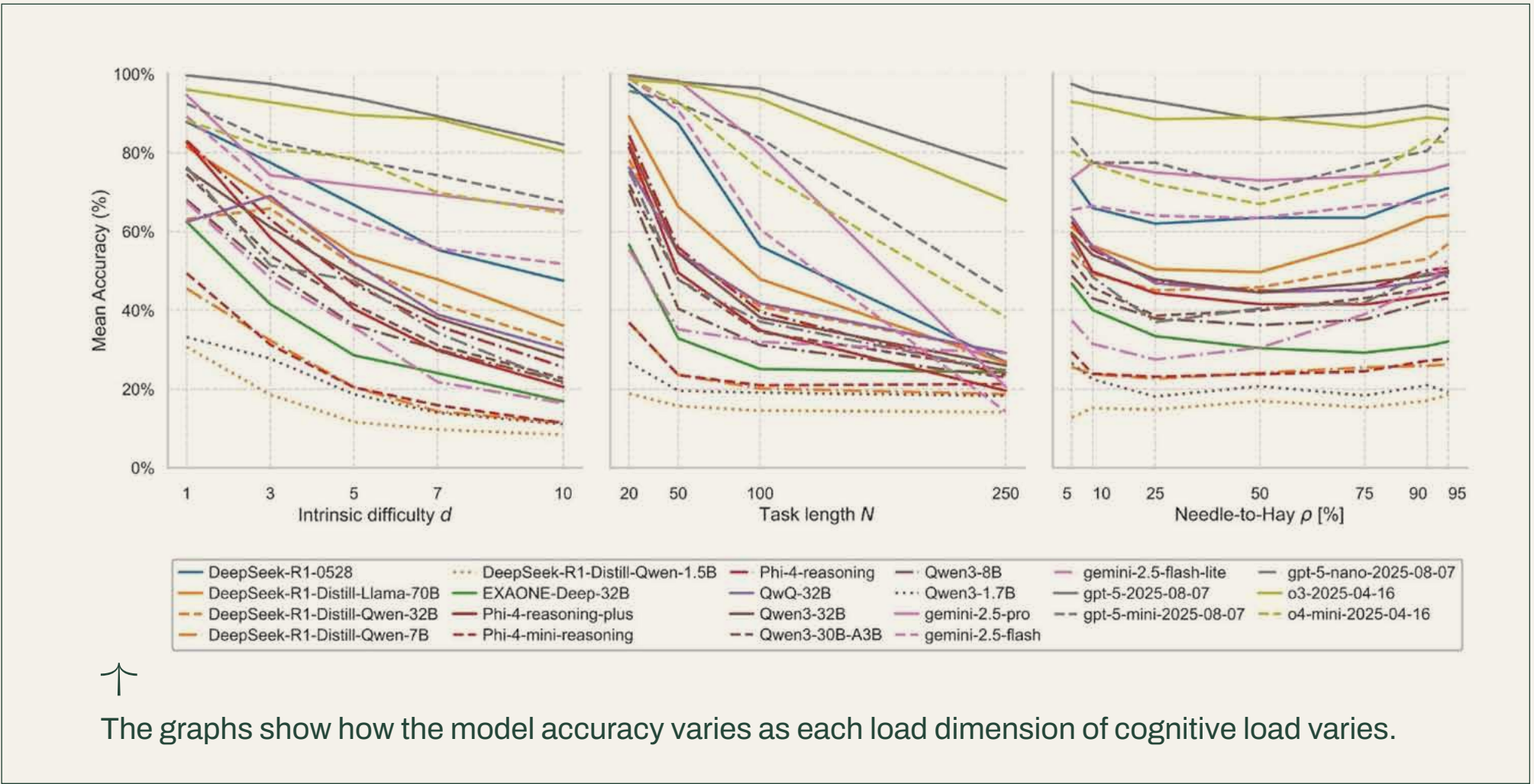
Kaiser tested CogniLoad on 22 open and commercial LLMs, including ChatGPT, DeepSeek, and Gemini. The results show that length, complexity, and irrelevant information significantly affect performance across models. Larger models generally handle longer and more complex reasoning chains better than smaller models. However, even the most advanced models show performance degradation as task demands increase.

The study demonstrates that current LLMs remain sensitive to structural properties of reasoning tasks. CogniLoad provides a controlled framework for identifying where and how reasoning breaks down, offering insights that are relevant for both model development and responsible deployment.

The work contributes to Integreat’s research on structure, scalability, and trustworthy machine learning by providing tools to better understand the limits of contemporary AI systems.

Kaiser, D., Frigessi, A., Ramezani-Kebrya, A., & Ricaud, B. CogniLoad:

A Synthetic Natural Language Reasoning Benchmark With Tunable Length, Intrinsic Difficulty, and Distractor Density. International Conference on Learning Representations (ICLR 2026).



Knowledge Acquisition and Description Logic

Understanding the Limits of Large Language Models

AAAI-26 outstanding papers

Model Change for Description Logic Concepts

Ana Ozaki, Jandson S Ribeiro

Integreat PI Associate Professor Ana Ozaki from the University of Oslo at the Department of Informatics, in the data and knowledge management group, worked together with colleagues and students on:

- (a) the development of algorithms for knowledge acquisition from language models and
- (b) on the theoretical study of model change operators for description logic concepts.

Description logic is a formalism for knowledge representation. Within this formalism, one can describe for example that “Minors cannot be drivers”, “Bees are insects that produce honey”, etc, with a mathematically defined semantics which allows for logic-based automated reasoning. The most well-known examples of description logic ontologies are for the biomedical sciences. These ontologies can be found in large repositories such as the Bio Portal. One of the main advantages of building ontologies based on logic is that one can specify the information, check which information can be inferred, and explain how it can be inferred, in a declarative, explicit, controlled, and verifiable manner. From a practical perspective, acquiring knowledge in the format of description logic ontologies can support the development and maintenance of ontologies for such repositories. Since building and maintaining such

ontologies manually is expensive, the enhanced capabilities of modern language models offer great opportunities for knowledge acquisition. In fact, the development of algorithms that interact with language models can partially automate the task of building ontologies. This has been a topic of study that Ana Ozaki presented in a keynote talk at the [KR 2025 conference](#) – main conference in Knowledge Representation (KR) – and in an [ECAI 2025 paper](#). The KR community has a strong focus on theoretical research, where researchers study how an agent that abstracts the behaviour of a system, robot, etc, can represent beliefs and update them when confronted with new information. Ana Ozaki together with her colleague Jandson S. Ribeiro investigated belief change operators for description logic concepts. This more theoretical work received a [best paper award at AAAI 2026](#).



Associate Professor Ana Ozaki

Abstract

We consider the problem of modifying a description logic concept in light of models represented as pointed interpretations. We call this setting model change, and distinguish three main kinds of changes: eviction, which consists of only removing models; reception, which incorporates models; and revision, which combines removal with incorporation of models in a single operation. We introduce a formal notion of revision and argue that it does not reduce to a simple combination of eviction and reception, contrary to intuition. We provide positive and negative results on the compatibility of eviction and reception for EL-bottom and ALC description logic concepts and on the compatibility of revision for ALC concepts.

Publications



Visit webpage

Aguas, J. de, Henckel, L., Pensar, J., & Biele, G. (2025). Causal Inference amid Missingness-Specific Independences and Mechanism Shifts. *Proceedings of the Forty-First Conference on Uncertainty in Artificial Intelligence*, 31–44.

<https://proceedings.mlr.press/v286/aguas25a.html>

Aguas, J. de, Krumscheid, S., Pensar, J., & Biele, G. (2025). The Probability of Tiered Benefit: Partial Identification with Robust and Stable Inference. *Proceedings of the Fourth Conference on Causal Learning and Reasoning*, 90–113.

<https://proceedings.mlr.press/v275/aguas25a.html>

Aguas, J. de, Pensar, J., Pérez, T. V., & Biele, G. (2025). Recovery and inference of causal effects with sequential adjustment for confounding and attrition. *Journal of Causal Inference*, 13(1).

<https://doi.org/10.1515/jci-2024-0009>

Asgari, F., Alamatsaz, M. H., Hayati, S., & Vitelli, V. (2025). Likelihood analysis of latent functional response regression models for sequences of correlated binary data. *Scandinavian Journal of Statistics*, 52(2), 840–872.

<https://doi.org/10.1111/sjos.12773>

Athanasopoulos, A., George, A.-M., & Dimitrakakis, C. (2025). Probably Correct Optimal Stable Matching for Two-Sided Market Under Uncertainty. In S. Das, A. Nowé, & Y. Vorobeychik (Eds), *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2025, Detroit, MI, USA, May 19-23, 2025* (pp. 152–160). International Foundation for Autonomous Agents and Multiagent Systems / ACM.

<https://doi.org/10.5555/3709347.3743527>

Bertinelli Salucci, C., Bakdi, A., Glad, I. K., Lindqvist, B. H., Vanem, E., & De Bin, R. (2025). Lifetime analysis with monotonic degradation: A boosted first hitting time model based on a homogeneous gamma process. *Lifetime Data Analysis*, 31(2), 300–339.

<https://doi.org/10.1007/s10985-025-09648-z>

Blum, M., Plenz, M., Ell, B., & Cimiano, P. (2025). Finding Good Neighbors: Examining the Importance of Neighborhood Selection for Link Prediction. In C. Shimizu, S. Ferrada, & L. Kagal (Eds), *Proceedings of the 13th Knowledge Capture Conference 2025, K-CAP 2025, Dayton, Ohio, USA, December 10-12, 2025* (pp. 90–93). ACM.

<https://doi.org/10.1145/3731443.3771352>

Choi, C., Subramaniam, A., Handegard, N. O., Ramezani-Kebrya, A., & Jenssen, R. (2025). Leveraging Foundation Model Adapters to Enable Robust and Semantic Underwater Exploration. In R. Jenssen & K. Bach (Eds), *Proceedings of the Symposium of the Norwegian AI Society 2025, Tromsø, Norway, June 17-18, 2025* (Vol. 3975, pp. 72–80). CEUR-WS.org.

<https://ceur-ws.org/Vol-3975/paper8.pdf>

Engebretsen, S., Thorarinsdottir, T. L., Diz-Lois Palomares, A., Storvik, G., Frigessi, A., & de Blasio, B. F. (2026). Solveig Engebretsen, Alfonso Diz-Lois Palomares, Birgitte Freiesleben de Blasio, Arnoldo Frigessi, Geir Storvik, and Thordis L. Thorarinsdottir's contribution to the Discussion of ‘Some statistical aspects of the Covid-19 response’ by Wood et al. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 189(1), 71–72.

<https://doi.org/10.1093/jrsssa/qnaf080>

Fedorova, M., Kutuzov, A., Periti, F., & Scherrer, Y. (2025). Explaining novel senses using definition generation with open language models. In C. Christodoulopoulos, T. Chakraborty, C. Rose, & V. Peng (Eds), *Findings of the Association for Computational Linguistics: EMNLP 2025* (pp. 22294–22302). Association for Computational Linguistics.

<https://doi.org/10.18653/v1/2025.findings-emnlp.1214>

Ferreira, C. E. A., Ell, B., Giese, M., & Carbonera, J. L. (2025). Investigating Dynamic Example Selection for ICL-Based Text-to-SPARQL Generation. *2025 IEEE 37th International Conference on Tools with Artificial Intelligence (ICTAI)*, 1412–1419.

<https://doi.org/10.1109/ICTAI66417.2025.00204>

Gehring, L., Blum, M., Ell, B., & Cimiano, P. (2025). Path of Time: Explanations for Temporal Knowledge Graph Completion through Chronological Regulation. *Joint Proceedings of the ESWC 2025 Workshops and Tutorials*.

<https://ceur-ws.org/Vol-3977/XAIKG-1.pdf>

George, A.-M., & Ozaki, A. (2025). On Middle Grounds for Preference Statements. *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2025, Montreal, Canada, August 16-22, 2025*, 4518–4525.

<https://doi.org/10.24963/IJCAI.2025/503>

Griesbauer, E., Czado, C., Frigessi, A., & Haff, I. H. (2025). TVineSynth: A Truncated C-Vine Copula Generator of Synthetic Tabular Data to Balance Privacy and Utility. *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, 3511–3519.

<https://proceedings.mlr.press/v258/griesbauer25a.html>

Grindheim, C. A., Stakkeland, M., Glad, I. K., & Vanem, E. (2025). Comparative Study of Machine Learning Methods for State of Health Estimation of Maritime Battery Systems. *Journal of Offshore Mechanics and Arctic Engineering*, 147(031404).

<https://doi.org/10.1115/1.4065967>

Hjort, A., Hermansen, G. H., Pensar, J., & Williams, J. P. (2025). Uncertainty quantification in automated valuation models with spatially weighted conformal prediction. *International Journal of Data Science and Analytics*, 20(8), 7089–7106.

<https://doi.org/10.1007/s41060-025-00862-4>

Kaiser, D., Frigessi, A., Ramezani-Kebrya, A., & Ricaud, B. *CogniLoad: A Synthetic Natural Language Reasoning Benchmark with Tunable Length, Intrinsic Difficulty, and Distractor Density*. International Conference on Learning Representations (ICLR 2026).

Kapar, J., Koenen, N., & Jullum, M. (2026). What’s Wrong with Your Synthetic Tabular Data? Using Explainable AI to Evaluate Generative Models. In R. Guidotti, U. Schmid, & L. Longo (Eds), *Explainable Artificial Intelligence* (pp. 19–43). Springer Nature Switzerland.

https://doi.org/10.1007/978-3-032-08327-2_2

Karlgren, J., Artemova, E., Bojar, O., Engels, M. I., Mikhailov, V., Sindelár, P., Velldal, E., & Øvrelid, L. (2025). Overview of ELOQUENT 2025: Shared Tasks for Evaluating Generative Language Model Quality. In J. Carrillo-de-Albornoz, A. G. S. de Herrera, J. Gonzalo, L. Plaza, J. Mothe, F. Piroi, P. Rosso, D. Spina, G. Faggioli, & N. Ferro (Eds), *Experimental IR Meets Multilinguality, Multimodality, and Interaction—16th International Conference of the CLEF Association, CLEF 2025, Madrid, Spain, September 9-12, 2025, Proceedings* (Vol. 16089, pp. 224–241). Springer.

https://doi.org/10.1007/978-3-032-04354-2_14

Karlgren, J., Artemova, E., Bojar, O., Mikhailov, V., Sahlgren, M., Velldal, E., & Øvrelid, L. (2025). ELOQUENT CLEF Shared Tasks for Evaluation of Generative Language Model Quality, 2025 Edition. In C. Hauff, C. Macdonald, D. Jannach, G. Kazai, F. M. Nardini, F. Pinelli, F. Silvestri, & N. Tonellotto (Eds), *Advances in Information Retrieval—47th European Conference on Information Retrieval, ECIR 2025, Lucca, Italy, April 6-10, 2025, Proceedings, Part V* (Vol. 15576, pp. 366–372). Springer.

https://doi.org/10.1007/978-3-031-88720-8_56

Kızılaslan, F., Michael Swanson, D., & Vitelli, V. (2025). A Weibull mixture cure frailty model for high-dimensional covariates. *Statistical Methods in Medical Research*, 34(6), 1192–1218.

<https://doi.org/10.1177/09622802251327687>

Kostylev, E. V. (2025). Foundations of Graph Neural Networks (A Logician’s View) (Invited Paper). In A. Artale, M. Bienvenu, Y. Ibáñez-García, & F. Murlak (Eds), *Joint Proceedings of the 20th and 21st Reasoning Web Summer Schools, RW 2024, Bucharest, Romania, September 19-22, 2024 / RW 2025, Istanbul, Turkey, September 25-28, 2025* (Vol. 138, p. 3:1-3:19). Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

<https://doi.org/10.4230/OASICS.RW.2024/2025.3>

Kvisgaard, V., & Pensar, J. (2025a). Bayesian Estimation of Causal Effects from Observational Categorical Data. *Bayesian Analysis*, 1(1), 1–28.

<https://doi.org/10.1214/25-BA1540>

Kvisgaard, V., & Pensar, J. (2025b). Bayesian IDA with Background Knowledge. In E. di Bella, V. Gioia, C. Lagazio, & S. Zaccarin (Eds), *Statistics for Innovation I* (pp. 91–96). Springer Nature Switzerland.

https://doi.org/10.1007/978-3-031-96736-8_16



Magnini, M., Squarcialupi, R., Sterri, M. T., & Ozaki, A. (2025). *Actively Learning EL Terminologies from Large Language Models*. ECAI 2025.
<https://doi.org/10.3233/faia251009>

Meggio, C., Pensar, J., Balcells, D., & De Bin, R. (2026). A path-based boosting algorithm for exploring transition metal compounds. *Statistical Modelling*, 1471082X251407458.
<https://doi.org/10.1177/1471082X251407458>

Mikhailov, V., Artemova, E., Butenko, Z., Øvrelid, L., & Velldal, E. (2025). Overview of the Preference Prediction Task at the ELOQUENT 2025 lab for evaluating generative language model quality. *Working Notes of CLEF*.

Mondorf, P., Wold, S., & Plank, B. (2025). Circuit Compositions: Exploring Modular Structures in Transformer-Based Language Models. In W. Che, J. Nabende, E. Shutova, & M. T. Pilehvar (Eds), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ACL 2025, Vienna, Austria, July 27—August 1, 2025 (pp. 14934–14955). Association for Computational Linguistics.
<https://aclanthology.org/2025.acl-long.727/>

Mosolygó, B., Opdahl, A. L., Xiao, G., Chen, J., & Ozaki, A. (2025). Benchmarking Knowledge Graph Question Answering via Complexity-Aware Queries. In J. Grabis & Y. Wautelet (Eds), *Advanced Information Systems Engineering Workshops* (pp. 30–37). Springer Nature Switzerland.
https://doi.org/10.1007/978-3-031-94931-9_3

Nguyen, A. D., Markov, I., Wu, F. Z., Ramezani-Kebrya, A., Antonakopoulos, K., Alistarh, D., & Cevher, V. (2025). Layer-wise Quantization for Quantized Optimistic Dual Averaging. *Forty-Second International Conference on Machine Learning, ICML 2025, Vancouver, BC, Canada, July 13-19, 2025*.
<https://openreview.net/forum?id=J6LYjEOxbz>

Olsen, L.H.B., Jullum, M. (2026). Improving the Weighting Strategy in KernelSHAP. In Guidotti, R., Schmid, U., Longo, L. (Eds), *Explainable Artificial Intelligence* (pp. 194–218). Springer Nature Switzerland.
https://doi.org/10.1007/978-3-032-08324-1_9

Ozaki, A. (2025). Actively Learning from Machine Learning Models with Queries and Counterexamples. In M. Console & B. Konev (Eds), *Reasoning Web. Declarative Artificial Intelligence: Knowledge, Rules, Logic: 19th International Summer School 2023 Oslo, Norway, September 21–24, 2023, Tutorial Lectures* (pp. 61–77). Springer Nature Switzerland.
https://doi.org/10.1007/978-3-031-80283-6_4

Ozaki, A., Confalonieri, R., Guimarães, R., & Imenes, A. (2025). Extracting PAC Decision Trees from Black Box Binary Classifiers: The Gender Bias Study Case on BERT-based Language Models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39, 19767–19775.
<https://doi.org/10.1609/aaai.v39i18.34177>

Scheel, I., & Gåsemyr, J. (2025). Graphical and Summary Diagnostics for Node Level Adequacy in Bayesian Hierarchical Models. *Bayesian Analysis*, 1(1), 1–26.
<https://doi.org/10.1214/24-BA1478>

Silva, T., Pedrini, H., & Rivera, A. R. (2025). Self-Organizing Visual Prototypes for Non-Parametric Representation Learning. *Forty-Second International Conference on Machine Learning, ICML 2025, Vancouver, BC, Canada, July 13-19, 2025*.
<https://openreview.net/forum?id=NGC7wdMFao>

Smajdor, A., Frigessi, A., & Zucknick, M. (2025). What is Right For ‘This’ Patient? Autonomy, Trust and Transparency in The Era of Medical Ai. *Síntese: Revista de Filosofia*, 52(164), 611–611.
<https://doi.org/10.20911/21769389v52n164p611/2025>

Varnet Pérez, T., Øvergaard, K. R., Frigessi, A., & Biele, G. (2025). Long-term effect of pharmacological treatment on academic achievement of Norwegian children diagnosed with ADHD: A target trial emulation. *International Journal of Epidemiology*, 54(2), dyaf010.
<https://doi.org/10.1093/ije/dyaf010>

Wally, Y., Mylius-Kroken, J., Kampffmeyer, M., Ehsani, R., Milosevic, V., & Wetzter, E. (2025, July 9). *Flatland and Beyond: Mutual Information Across Geometries*. 2nd Beyond Euclidean Workshop: Hyperbolic and Hyperspherical Learning for Computer Vision.
<https://openreview.net/forum?id=LX9bskkzH2#discussion>

Wold, S., Charpentier, L. G. G., & Simon, É. (2025). Systematic Generalization in Language Models Scales with Information Entropy. In W. Che, J. Nabende, E. Shutova, & M. T. Pilehvar (Eds), *Findings of the Association for Computational Linguistics, ACL 2025, Vienna, Austria, July 27—August 1, 2025* (pp. 1807–1819). Association for Computational Linguistics.
<https://aclanthology.org/2025.findings-acl.90/>

Wu, Z., Choi, C., Cao, X., Cevher, V., & Ramezani-Kebrya, A. (2025). Addressing Label Shift in Distributed Learning via Entropy Regularization. *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*.
<https://openreview.net/forum?id=kuYxecnlv2>

Zhou, D., Kharlamov, E., & Kostylev, E. V. (2025). GLoRa: A Benchmark to Evaluate the Ability to Learn Long-Range Dependencies in Graphs. *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*.
<https://openreview.net/forum?id=2jf5x5XoYk>

Zhou, Y., Zhou, B., Li, H., Lyu, Q., Qu, Y., Waaler, A., & Yu, I. C. (2025). Dataset for Industrial Question Answering with Explanation and Scalable Ensemble Generation. In G. Long, M. Blumestein, Y. Chang, L. Lewin-Eytan, Z. H. Huang, & E. Yom-Tov (Eds), *Companion Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025—2 May 2025* (pp. 825–828). ACM.
<https://doi.org/10.1145/3701716.3715310>

Zhou, Y., Zhou, B., & Yu, I. C. (2025). NMN-BART: Generating Natural Language Explanations for Visual Question Answering. In R. Jenssen & K. Bach (Eds), *Proceedings of the Symposium of the Norwegian AI Society 2025, Tromsø, Norway, June 17-18, 2025* (Vol. 3975, pp. 38–48). CEUR-WS.org.
<https://ceur-ws.org/Vol-3975/paper4.pdf>



Preprints:

Aasan, M., Hjelkrem-Tan, M., Catalano, N., Choi, C., & Rivera, A. R. (2025). *Differentiable Hierarchical Visual Tokenization*.
<https://doi.org/10.48550/ARXIV.2511.02652>

Arteaga, G. Y., Aasan, M., Chakraborty, R., Hjelkrem-Tan, M., Silva, T., Kampffmeyer, M., & Rivera, A. R. (2025). *Why Prototypes Collapse: Diagnosing and Preventing Partial Collapse in Prototypical Self-Supervised Learning*.
<https://doi.org/10.48550/ARXIV.2510.20108>

Eilertsen, B., Bjørgfinsdóttir, R., Vargas, F., & Ramezani-Kebrya, A. (2025). *Aligning Attention with Human Rationales for Self-Explaining Hate Speech Detection* (arXiv:2511.07065). arXiv.
<https://doi.org/10.48550/arXiv.2511.07065>

Ell, B. (2025). *Graph Pattern-based Association Rules Evaluated Under No-repeated-anything Semantics in the Graph Transactional Setting* (Version 1). arXiv.
<https://doi.org/10.48550/ARXIV.2512.15308>

Feest, J., Ruffieux, H., Lingjærde, C., & Xi, X. (2025). *NExON-Bayes: A Bayesian approach to network estimation informed by ordinal covariates* (arXiv:2508.21649). arXiv.
<https://doi.org/10.48550/arXiv.2508.21649>

Frommlet, F., Lachmann, J., Storvik, G., & Hubin, A. (2025). *FBMS: An R Package for Flexible Bayesian Model Selection and Model Averaging* (arXiv:2509.00753). arXiv.
<https://doi.org/10.48550/arXiv.2509.00753>

George, F., Khalid, M., Khan, A., Aslansefat, K., & Rivera, A. R. (2025). *Evaluating the Effectiveness of Diffusion Models in Cross-Domain Data Augmentation*. SSRN.
<https://doi.org/10.2139/ssrn.5087057>

Hjelkrem-Tan, M., Aasan, M., Arteaga, G. Y., & Rivera, A. R. (2025). *SPoT: Subpixel Placement of Tokens in Vision Transformers*.
<https://doi.org/10.48550/ARXIV.2507.01654>

Jahanmard, V., Hordoir, R., Ramezani-Kebrya, A., Ellmann, A., & Delpeche-Ellmann, N. (2025). *Principled Fourier Neural Operators for High-resolution Regional Ocean Modeling*. Preprints.
<https://doi.org/10.22541/essoar.176279300.07801097/v1>

Jahanmard, V., Ramezani-Kebrya, A., & Hordoir, R. (2025). *Principled Operator Learning in Ocean Dynamics: The Role of Temporal Structure* (arXiv:2510.09792). arXiv.
<https://doi.org/10.48550/arXiv.2510.09792>

Jullum, M., Olsen, L. H. B., Lachmann, J., & Redelmeier, A. (2025). *shapr: Explaining Machine Learning Models with Conditional Shapley Values in R and Python* (arXiv:2504.01842). arXiv.
<https://doi.org/10.48550/arXiv.2504.01842>

Kaiser, D., Frigessi, A., Ramezani-Kebrya, A., & Ricaud, B. (2025). *CogniLoad: A Synthetic Natural Language Reasoning Benchmark With Tunable Length, Intrinsic Difficulty, and Distractor Density* (arXiv:2509.18458). arXiv.
<https://doi.org/10.48550/arXiv.2509.18458>

Kızılaslan, F., & Vitelli, V. (2025). *Bayesian Semiparametric Mixture Cure (Frailty) Models* (arXiv:2512.08173). arXiv.
<https://doi.org/10.48550/arXiv.2512.08173>

Lee, A., Stoltenberg, E. A., & Mykland, P. A. (2025). *Semiparametrics via parametrics and contiguity* (arXiv:2501.09483). arXiv.
<https://doi.org/10.48550/arXiv.2501.09483>

Moen, P. A. J. (2025). *A general methodology for fast online changepoint detection* (arXiv:2504.09573). arXiv.
<https://doi.org/10.48550/arXiv.2504.09573>

Neglia, R., Cini, A., Bronstein, M. M., & Bianchi, F. M. (2025). *ResCP: Reservoir Conformal Prediction for Time Series Forecasting* (arXiv:2510.05060). arXiv.
<https://doi.org/10.48550/arXiv.2510.05060>

Rooijakkers, J., Rønneberg, L., Briol, F.-X., Knoblauch, J., & Altamirano, M. (2025). *Multi-Output Robust and Conjugate Gaussian Processes* (arXiv:2510.26401). arXiv.
<https://doi.org/10.48550/arXiv.2510.26401>

Salucci, C. B., Kvellestad, A., & Bin, R. D. (2025). *A note on the asymptotic distribution of the Likelihood Ratio Test statistic under boundary conditions* (arXiv:2509.00223). arXiv.
<https://doi.org/10.48550/arXiv.2509.00223>

Samuel, D., Øvreliid, L., Velldal, E., & Kutuzov, A. (2025). *Fluent Alignment with Disfluent Judges: Post-training for Lower-resource Languages* (arXiv:2512.08777). arXiv.
<https://doi.org/10.48550/arXiv.2512.08777>

Zobolas, J., George, A.-M., López, A., Fischer, S., Becker, M., & Aittokallio, T. (2025). *Optimizing Prognostic Biomarker Discovery in Pancreatic Cancer Through Hybrid Ensemble Feature Selection and Multi-Omics Data*.
<https://doi.org/10.48550/ARXIV.2509.02648>

International collaborations and mobility

Integreat researchers have extensive international collaboration, reflected in our list of publications. In 2025, Integreat hosted several international visiting researchers for short-term stays, organised guest lectures and international workshops, and participated in advisory boards, took on roles in international conferences, and engaged in seminar exchanges with partner institutions. Several Integreat researchers acted as area chair in the main international conferences, recognising the excellence of our staff.

We initialised developing a structured approach to mobility and visiting fellowships for the upcoming years; for example, early in 2026 Integreat will publish an open call for longer research stays, including PhD-level mobility.

Academic visitors at Integreat in 2025 (selection):

- **Professor Eyke Hüllermeier**, Institute of Informatics, LMU Munich, Germany
- **Professor Henrik Boström**, KTH Royal Institute of Technology, Sweden
- **Associate Professor Jeremias Knoblauch**, UCL's Department of Statistical Science, London, UK
- **Professor Steven Marron**, Department of Statistics and O.R., School of Data Science and Society, University of North Carolina, USA
- **Professor Willem Waegeman**, Department of Data Analysis and Mathematical Modelling, Ghent University, Belgium
- **Senior Lecturer David Hofmeyr**, STOR-i Centre for Doctoral Training, UK



Early-career researchers and students (selection):

- **PhD Fellow Ana Niño López**, University of Cádiz, Spain
- **Lecturer Carolina Euan**, STOR-i Centre for Doctoral Training, UK
- **Lecturer Israel Martinez-Hernandez**, STOR-i Centre for Doctoral Training, UK
- **PhD Fellow Kristina Grolmusova**, STOR-i Centre for Doctoral Training, UK
- **PhD Fellow Kristina Bratkova**, STOR-i Centre for Doctoral Training, UK
- **PhD Fellow María Eugenia Velasco García**, University of Granada, Spain
- **MA student Sara Spieler**, Université Côte d'Azur in Nice, France
- **PhD Fellow Theo Crookes**, STOR-i Centre for Doctoral Training, UK

Integreat has partnerships with:

- **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 Pls supervise three PhD students at STORi, who spend short periods in Oslo. Glad is a member of the External Advisory Committee of STORi, and Idris Eckley, director of STORi, is visiting professor at Integreat.



Visiting Integreat has left an enduring and positive memory with me: the institute's staff made me welcome every step of the way, and I went home with a lot of new interesting research directions to investigate. In fact, my visit led directly to a collaboration with one of the academics based at the institute and has already elicited a published paper and reciprocal visit to London.

— Visiting researcher Jeremias Knoblauch

International workshops and summer schools

Integreat organised two workshops/summer schools, dedicated to Veridical AI and preference learning in 2025.

Veridical AI was co-organised by Integreat, the Department of Statistics, University of California, Berkeley, the L'EMbeDS Department of Excellence at the Sant'Anna School of Advanced Studies at Pisa, the Department of Methods and Models for Economics, Territory and Finance of the University of Roma La Sapienza and the Institute for Systems Analysis and Computer Science of the Italian National Research Council. The workshop showcased and promoted veridical (truthful) data science towards reproducible and reliable data analysis and decisions. It intends to build a community of researchers in Europe for trustworthy data science and machine learning. The next workshop will be in 2026 in Paris, France.

Integreat organised the second International Summer School on Preference Learning (PrefStat). PrefStat is a permanent inter-university research and working group that aims at disseminating statistical approaches in

Preference Learning, understood as modelling and analysis of preference rankings and ordinal data related to preference analysis. The course was designed to provide a comprehensive overview of preference statistics, a rapidly growing field that has gained significant attention in recent years due to the numerous application fields: recommender systems, Large Language Models, surveys and psychological experiments, marketing, economics, and political science. Preference learning is concerned with all data analyses involving preferences, rankings, ratings, clicking, or any kind of ordinal data. The school provided a deep introduction to the topic and insight into more challenging tasks that are of interest in modern applications, such as handling partial, unstructured, exogenous information, individual preference prediction, and importance feature selection. PrefStat 2026 will happen in Valladolid, Spain.



Integreat working groups, reading clubs 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?previous_action=info

Seminars last 45 minutes and are followed by discussion. All sessions are streamed.



[Visit webpage](https://sympa.uio.no/integreat.no/subscribe/seminars?previous_action=info)

- **Integreat Wednesday’s Seminars “allmøte”**

Every Wednesday, the Integreat community gathers for a short internal seminar followed by discussion and information sharing. Topics include recently submitted papers, conference reports, and introductions of new members and visitors. A simple lunch is provided. The meeting is streamed between Oslo and Tromsø to support integration across hubs.

- **Thursday Tromsø–Oslo Coffee**

A weekly 30-minute informal digital meeting connecting the two hubs. This low-threshold format strengthens social ties and facilitates informal academic exchange.

Reading Groups

Reading groups are dynamic and often initiated organically in response to shared research interests. Some are open to all interested participants, while others are linked to specific projects and organised by invitation. Current and recent groups include:

- AI and Philosophy
- Predictive Bayes and post Bayes
- Learning from human preferences
- Inconsistency Learning Club
- Reading group on Bayesian neural networks
- Reading group on data-driven tensor factorization models
- XAI + LLM reading group

Thematic Research Seminars

Several research groups organise their own thematic seminar series:

- LTG Research Seminar
- Tuesday Seminar in Statistics and Data Science
- OCBE Biostatistical Seminars
- Visual Intelligence Seminars

Workshop on AI and Philosophy

In December 2025, Integreat hosted a workshop on AI and Philosophy at the University of Oslo, bringing together researchers from philosophy, informatics, mathematics, and related disciplines. The workshop created space for dialogue across disciplinary traditions, with short flash talks from senior researchers, early-career scholars, and students.

The aim was to present ongoing work and to explore shared questions at the intersection of artificial intelligence, logic, ethics, and conceptual analysis. Discussions addressed how philosophical reflection can inform technical development, and how formal methods in AI challenge and reshape philosophical thinking.

The workshop also served as a starting point for building a more sustained arena for exchange, with interest in establishing an ongoing reading or working group. The event reflected Integreat's commitment to working with disciplinary differences, and to treating conceptual clarity as integral to scientific progress.



Integreat voices

Research projects at Integreat are designed across research themes, objectives, and research traditions. Our research questions include an interdisciplinary team of PIs and it is rarely limited to one methodological or conceptual frame. How does interdisciplinary approach unfold in everyday life, supervision and collaboration. Is it straightforward and easy? From our experiences, we see that it is a matter of daily work — of supervision meetings, methodological choices, and learning to articulate assumptions that are often left implicit.

Each year, we highlight one perspective from within the centre. Last year, we presented reflections on the working day of a PhD candidate. This year, we turn to interdisciplinarity in practice. The reflections below offer two viewpoints on a doctoral project supervised across five disciplines — illustrating how complexity, when structured and trusted, becomes a source of intellectual strength rather than fragmentation.

Handling Inconsistencies: a Joint Statistical Learning and Logical Reasoning Framework

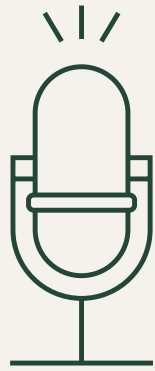
This project looks at how different fields deal with conflicting and inconsistent information. By connecting ideas from statistics and knowledge representation, we work toward a common language that helps bridge probabilistic approaches and logical reasoning, encouraging interdisciplinary collaboration.

Project team

- **Lorenzo Zuccato** (PhD)
- **Supervisors:** **Ida Scheel** (Main Supervisor), University of Oslo, Department of Mathematics
- **Arnoldo Frigessi**, University of Oslo, Department of Mathematics
- **Anne-Marie George**, University of Oslo, Department of Informatics
- **Martin Giese**, University of Oslo, Department of Informatics
- **Valeria Vitelli**, University of Oslo, Department of Biostatistics

Research themes:

- RT1, RT2



PhD voice: Lorenzo Zuccato

– Success through learning
to navigate difference



Lorenzo ↑

“When I started my PhD, I thought having five supervisors would mean constant guidance. What it actually meant was learning to navigate five ways of thinking — sometimes aligned, sometimes not. Early on, this was disorienting, and there were moments when I struggled to know whose advice to follow.

What helped was learning to make my own reasoning explicit and to use supervision

meetings as spaces for clarification rather than resolution. Over time, I became more confident in holding different perspectives at once and in making informed choices between them. I believe that by the end of the PhD, this will become a strength: I will no longer wait for a single ‘right’ answer, but able to work productively with complexity. That has shaped both my research and how I see my role as a researcher going forward.”

Supervisors' perspective

– Success through structure and trust

“We sometimes say that interdisciplinarity begins at the level of ideas. In practice, it begins with how people choose to work together. One of our PhD candidates is supervised by five researchers from different disciplines. This has required time, patience, and a willingness to explain disciplinary assumptions that are often left implicit.

What made this supervision model work was a shared understanding that disagreement was part of the process, not a problem to be eliminated. Roles and expectations were clarified early, and space was created for discussion rather than quick convergence. The result has been a strong doctoral project and a researcher trained to navigate complexity with confidence — which, for us, is a clear marker of success. We also feel that we learn a lot more in this process than in a 'standard' PhD project, so it also strengthens the quality of the centre at the senior level” – say the supervisors Professor Martin Giese, Associate Professor Ida Scheel, and Associate Professor Anne-Marie George.



Supervision autumn 2025: Lorenzo with the team of supervisors

Philosopher's voice

Strengthening Ethics in Practice: Recruiting Across Disciplines



Yael ↑

In 2025, Integreat made a strategic decision to strengthen the Ethics research theme. Rather than treating ethical reflection as an external commentary on technical work, the centre chose to embed philosophical expertise directly within its research environment. This led to the recruitment of a philosopher to a position at the Department of Mathematics.

The decision raised practical and intellectual questions from the outset. What does it mean to employ across disciplines when academic career paths, publication cultures, and evaluation criteria differ? How should career development be structured when a researcher does not sit neatly within one disciplinary tradition? Is there a risk of “falling between two chairs,” or does working between disciplines produce a stronger and more independent profile?

The recruitment also tested the everyday realities of interdisciplinary collaboration. How does philosophical analysis meet mathematical modelling? Do researchers share a language, or must it be developed? What becomes easier — and what becomes more demanding — when disciplinary assumptions are made explicit?

The reflections that follow address these questions from the perspective of the newly recruited philosopher. They describe what it means to enter a mathematics and natural science environment as a philosopher: what was surprising, what was challenging, what proved unexpectedly productive, and what advice might be offered to future interdisciplinary researchers.

I applied for the postdoctoral position in spring 2025 and started working at Integreat in September 2025.



Real-life situation: Yael explaining her research.

After five years at UiO, I thought I was familiar with the university's culture, but it gradually became clear that each research environment has its own way of working.

I knew that researchers at Integreat collaborate closely, but I had not anticipated how much this would shape everyday routines. I came from a work rhythm that was largely individual; at Integreat, I stepped into shared *Overleaf* documents, open office discussions, and daily communal lunches.

Over the first months, I have found that, despite differences in disciplinary background and language, we share similar concerns and ambitions for how our research should be conducted. There is still work to be done in building common conceptual ground and shared terminology, but I have experienced an openness to

dialogue and experimentation, where philosophy and ethics are not add-ons but integrated parts of project development.

“I am looking forward to seeing how this collaboration continues to develop and what we will achieve together.”

Integreat events



Seminars 2025

Seminars
2025



TUESDAY SEMINAR 14.10

Tommy Sonne Alstrøm

Explainable AI for time series
– a signal processing approach



TUESDAY SEMINAR 10.06

Frank van Harmelen

Systems that learn and reason



TUESDAY SEMINAR 29.04

Willem Waegeman

Making AI systems more
trustworthy through uncertainty
disentanglement



TUTORIAL 26.02

Eyke Hüllermeier

Uncertainty in machine learning:
from aleatoric to epistemic



TUESDAY SEMINAR 30.09

Gavin Taylor

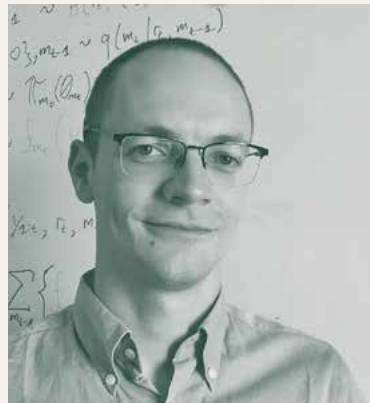
Evading AI systems



AI WITH PRIDE 04.06

Elisabeth Wetzer

Exploring inclusivity and
representation in artificial
intelligence



TUESDAY SEMINAR 01.04

Jeremias Knoblauch

Post-Bayesian Machine Learning



GUEST TALK 06.02

Marvin Wright

Tree-based generative
modeling with applications
in explainable AI



TUESDAY SEMINAR 16.12

Francesco Locatello

Causal inference in high-
dimensional exploratory
experiments



TUESDAY SEMINAR 23.09

Sophie Belman

The role of the environment
in the dynamics and evolution
of microbial populations



GUEST TALK 28.05

Steve Marron

Data Integration Via analysis
of Subspaces (DIVAS)



TUESDAY SEMINAR 18.03

Artur d'Avila Garcez

Neurosymbolic AI



TUESDAY SEMINAR 04.02

Robinson Hordoir

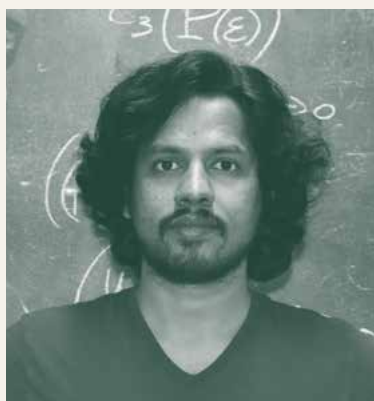
Barotropic trends through Barents
Sea opening for the period
1975–2021: The underlying
physical processes explained
using Artificial Intelligence



TUESDAY SEMINAR 11.11

Sampo Pyysalo

Register always matters: analysis
of LLM pretraining data through
the lens of language variation



TUESDAY SEMINAR 16.09

Senthil Murugan
Nagarajan

From data to decisions:
knowledge-driven ML
in orthopedic surgery



TUESDAY SEMINAR 27.05

Henri Prade

Reasoning and learning
in the setting ory – Overview
and perspectives



GUEST TALK 05.03

Mona Singh

Language models to characterize
protein sequences: their power,
challenges, and future directions



TUESDAY SEMINAR 21.01

LaVonne Roberts,
Eric Kern

Restoring the Self: Personal
LLMs as guardians of human
identity and agency for people
living with severe communica-
tion disorders



TUESDAY SEMINAR 28.10

Namhoon Lee

The unseen frontier: pushing
the limits of LLM sparsity with
surrogate-free ADMM



SPECIAL SEMINAR 25.08

Tengyao Wang

Coverage correlation: detecting
singular dependencies between
random variables



TUESDAY SEMINAR 13.05

Anis Yazidi

On RL without Lock in Action



GUEST TALK 27.02

Henrik Boström

Uncertainty quantification
through conformal regressors
and predictive systems

The Second Annual Retreat: From gathering to working together



In November 2025, the Integreat community gathered in Oslo for the second annual gathering. Over three days, researchers from across themes and partners met at the University of Oslo to present ongoing work, discuss methodological challenges, and reflect on the centre’s development.

25–27 November,
→ Oslo

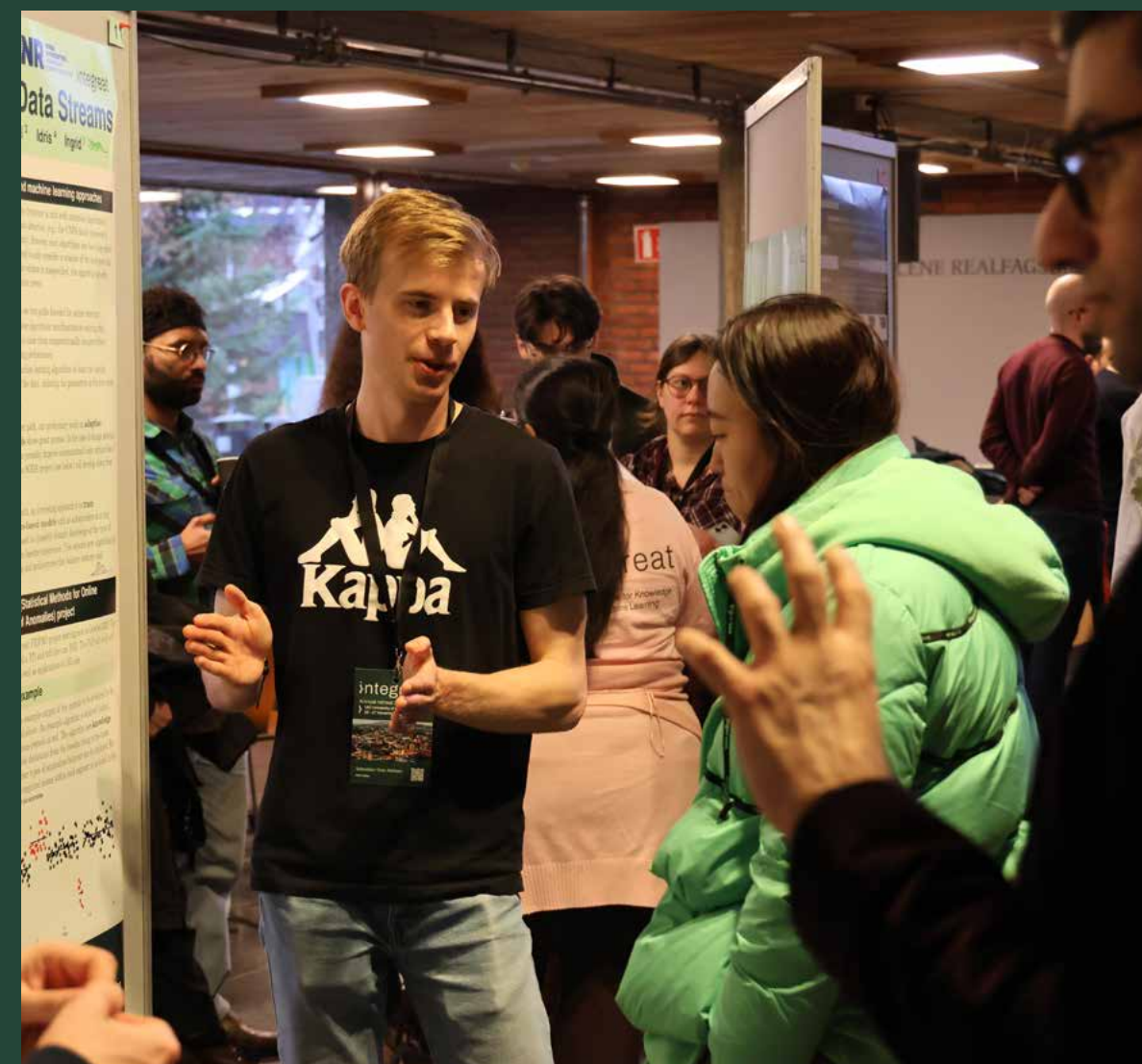
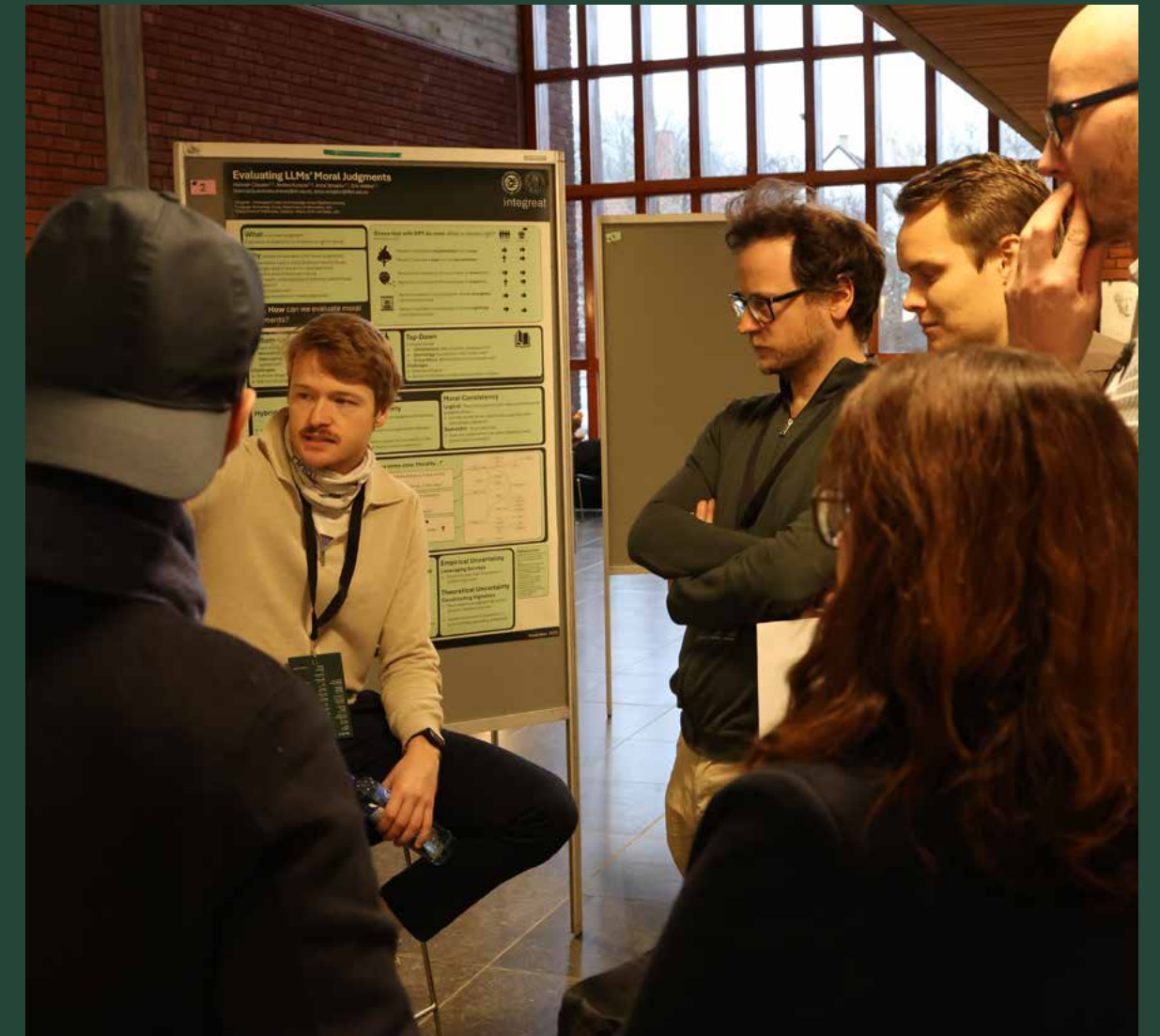
If the first retreat marked the beginning of a shared journey, the second revealed how the collaboration has matured. The newly introduced concept of a Poster Safari became a central feature of the retreat in November. Physically moving between projects, each presented in a poster, researchers engaged directly with one another’s work, asking questions and identifying connections across themes. In addition, participants were divided in small groups, with 4-5 members, and each group had a specific task related to characterising the collection of projects: what knowledge are we “integreating”? Status on the sustainable AI? What does interdisciplinary means at Integreat? And so on. The day after, each group reported their findings to the Integreat community.

The retreat also invited reflection beyond individual projects. A panel discussion on impacts of AI, moderated by Integreat codirector Benjamin Ricaud, brought together very interesting perspectives from a secondary school student, an artist, and a representative from the Research Council of Norway. Their contributions invited researchers to consider how AI is experienced outside research and how trust, usability, and public understanding intersect with technical development.

Questions of research practice were addressed by the invited speaker Lonni Besançon, who spoke about scientific transparency and challenges within contemporary publication systems. His talk prompted open discussion about responsibility, integrity, and the norms that shape scientific work.

Before the annual retreat, PhD candidates and postdoctoral fellows gathered for a dedicated INTREF pre-day. This session provided space for early-career researchers to meet across projects, exchange experiences, and strengthen their own network within the centre. Programme included a workshop on elevator pitching, equipping participants with tools to communicate their research clearly and confidently across disciplines and audiences. The workshop was followed by peer presentations of posters, allowing researchers to test ideas, receive feedback, and refine how they articulate their projects.

The retreat also marked a transition in leadership, with Ingrid Glad announced as the incoming Director of Integreat, while Arnaldo Frigessi leaving the post to continue as Co-director and leader of the new centre TRUST – The Norwegian Centre for Trustworthy AI.





The Integreat Safari concept

The Integreat Safari is a structured exploration of the centre’s research landscape. During the annual retreat, projects are presented across three rounds of poster sessions, with participants moving in small, mixed groups each group guided by a specific question.

The format encourages researchers to look beyond their own themes, identify connections, and reflect on the centre’s broader direction.

Tasks that the Safari groups were given:

Groups 1 and 2

TASK — Integreat has four Objectives (O1 Accurate and Knowledge-driven ML, O2 Sustainable and Green ML, O3 Fair, Explainable and Trustworthy ML, and O4 ML with Quantified Uncertainty). Please find out how the objective on Fair, Explainable and Trustworthy ML is covered in the project portfolio. Please look carefully, not only spotting the big elephants, but also the very smallest animals. Please be open minded and remember that fairness and explainability might come in various forms. For more info on Integreat Objectives, consult <https://www.integreat.no/research/>

FORMAT — Please present your findings in form of a panel discussion, with one moderator, and 2–3 participants from each of the two groups. The panel discussion should be informative and easy to follow. Possible critique should be directed towards the total project portfolio, not individual projects.

Groups 3 and 4

TASK — Integreat has four Objectives (O1 Accurate and Knowledge-driven ML, O2 Sustainable and Green ML, O3 Fair, Explainable and Trustworthy ML, and O4 ML with Quantified Uncertainty). Please find out how the objective on Sustainable and Green ML is covered in the project portfolio. Please look carefully, not only spotting big elephants, but also the very smallest animals. Please be open minded and remember that sustainability might come in various forms. For more info on Integreat Objectives, consult <https://www.integreat.no/research/>

FORMAT — Please present your findings in form of a panel discussion, with one moderator, and 2–3 participants from each of the two groups. The panel discussion should be informative and easy to follow. Possible critique should be directed towards the total project portfolio, not individual projects.

Groups 5 and 6

TASK — Integreat has four Objectives (O1 Accurate and Knowledge-driven ML, O2 Sustainable and Green ML, O3 Fair, Explainable and Trustworthy ML, and O4 ML with Quantified Uncertainty). Please find out how the objective on uncertainty quantification is covered in the project portfolio. Please look carefully, not only spotting elephants, but also the very smallest animals. Please be open minded and remember that uncertainty quantification might come in various forms. For more info on Integreat Objectives, consult <https://www.integreat.no/research/>

FORMAT — Please present your findings in form of a panel discussion, with one moderator, and 2-3 participants from each of the two groups. The panel discussion should be informative and easy to follow. Possible critique should be directed towards the total project portfolio, not individual projects.



Groups 7, 8 and 9

TASK — Integreat = Norwegian Centre for Knowledge-driven Machine Learning. Please study the project portfolio to uncover what types of knowledge we are using, and the various ways such knowledge is integrated in algorithms, methods and/or theory.

FORMAT — Group 7 will start the presentation with a short overview, for example sorting the different types of knowledge spotted into natural clusters. Group 8 and 9 will follow up, with more in depth analysis, dividing the material between them. The three groups have to coordinate well. Oral presentation by group representatives, with or without slides.

Groups 10 and 11

TASK — Please go hunting for real data and possible applications present in the posters. What kind of datasets and which real world (science, industry, societal, health, etc.) problems are we working on? How strongly are the applications motivating the research questions?

FORMAT — Please present your findings in a presentation with slides – 10 minutes from each group. Remember to include photos of some of the animals you have spotted.

Groups 12 and 13

TASK: Many of our projects have known and unknown similarities and connections. Please suggest high level labels on natural clusters of projects, where projects might be in multiple clusters. Also, look out for animals that should have been in our portfolio, but that have escaped.

FORMAT: The groups should perform an on stage dialog between the two groups, presenting and discussing their findings. The two groups should stand on each side of the stage, and the discussion should go back and forth in a natural, unstructured way. The groups can bring one slide each.

Group 14

TASK: Congratulations, you are the artistic group. You will be provided with pdfs of all posters. Use LLMs or other AI tools to make a rap/poem/song about the Integreat Safari posters.

FORMAT: Please make an artistic performance of your choice, max 5 minutes.

→ This illustration of Integreat is based on a visual concept developed by Group 12. The illustration is later refined by the designer.



Imagine – Integreat

Group 14 translated the Integreat Safari posters into a short AI-supported song set to the melody of Imagine, offering a creative interpretation of the centre's interdisciplinary research landscape.

Imagine no divisions,
In logic and ML,
Ethics and statistics,
Together standing well
Imagine all the breakthroughs
Accepted in NeurIPS ...

Imagine no reviewers,
It's easy if you try,
No walls between fields,
Just knowledge flowing by.
Imagine all the scientists
Sharing what they know ...

You may say I'm a dreamer,
But I'm not the only one.
I am happy that you joined us
And Integreat becoming one

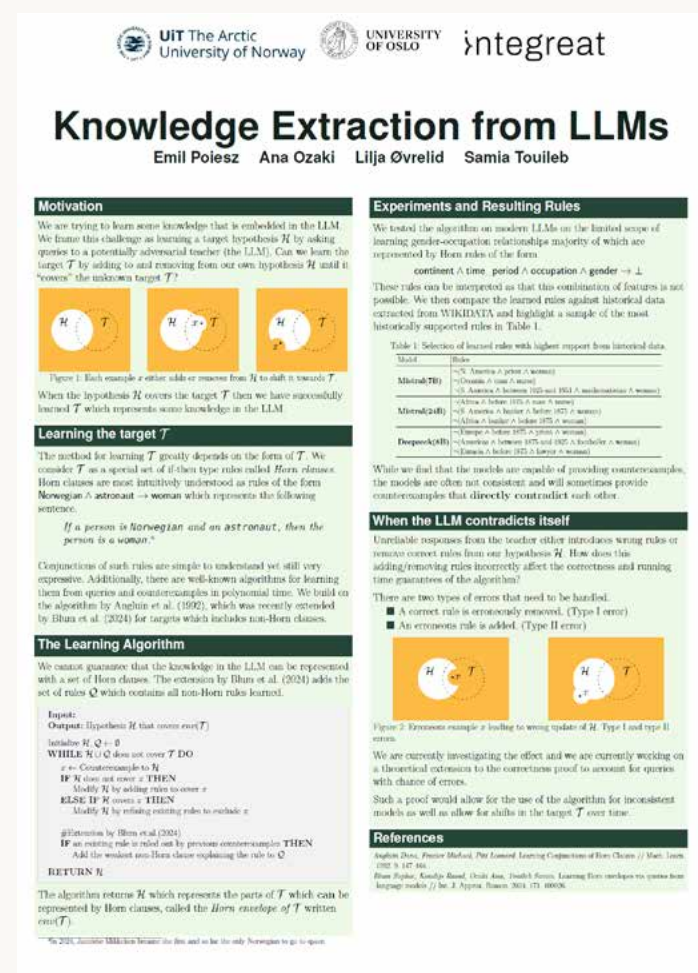
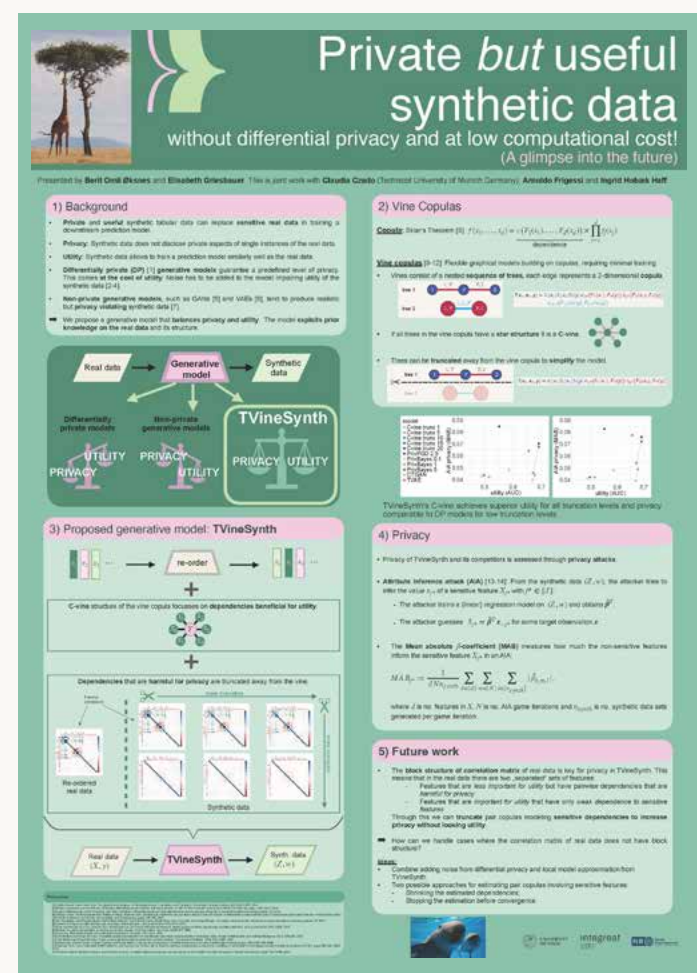
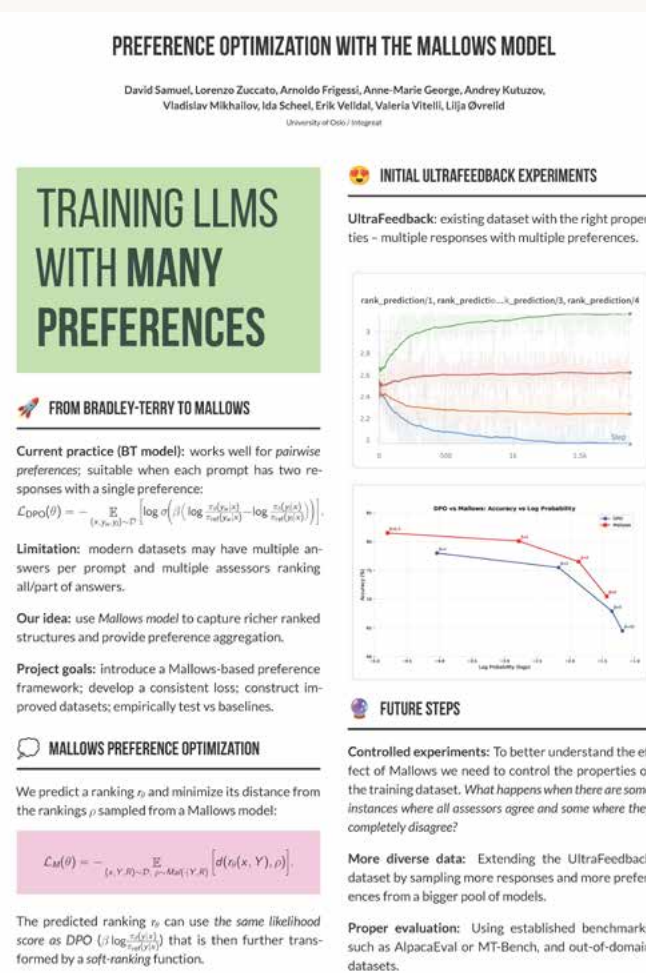
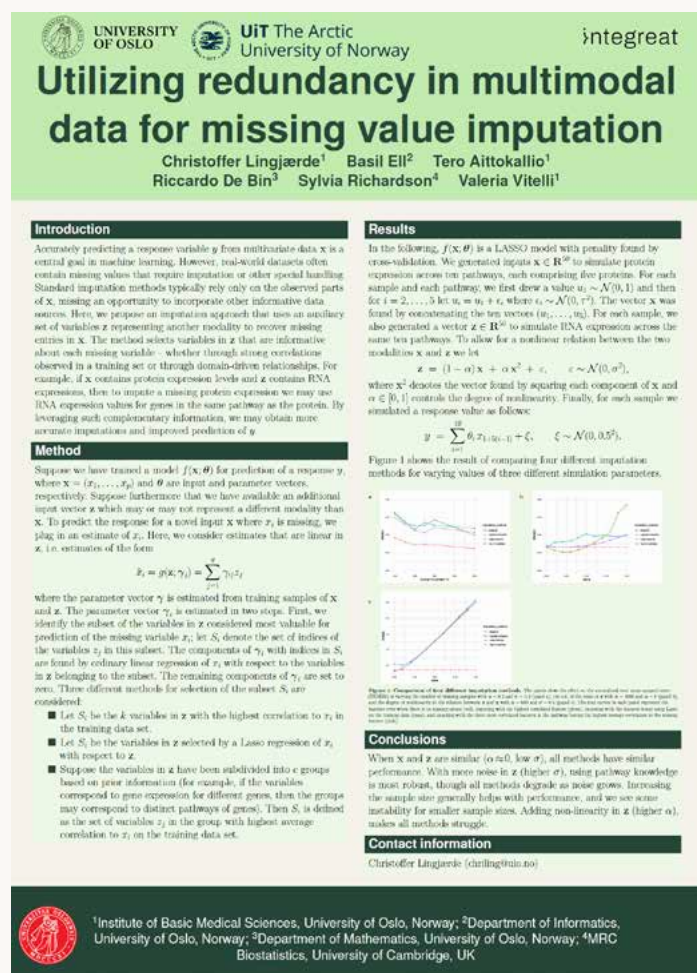
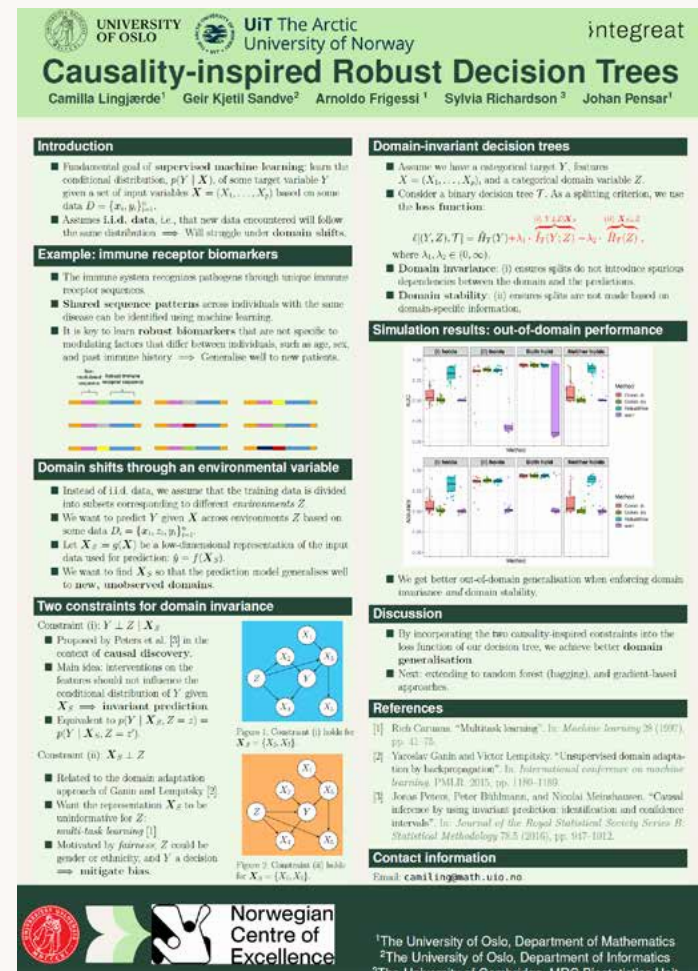
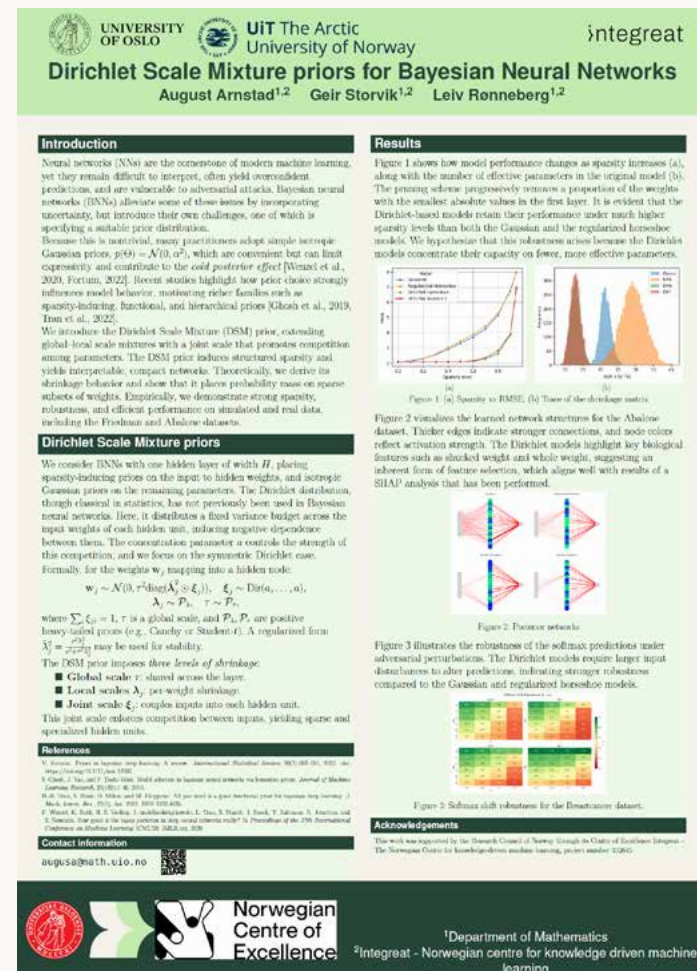
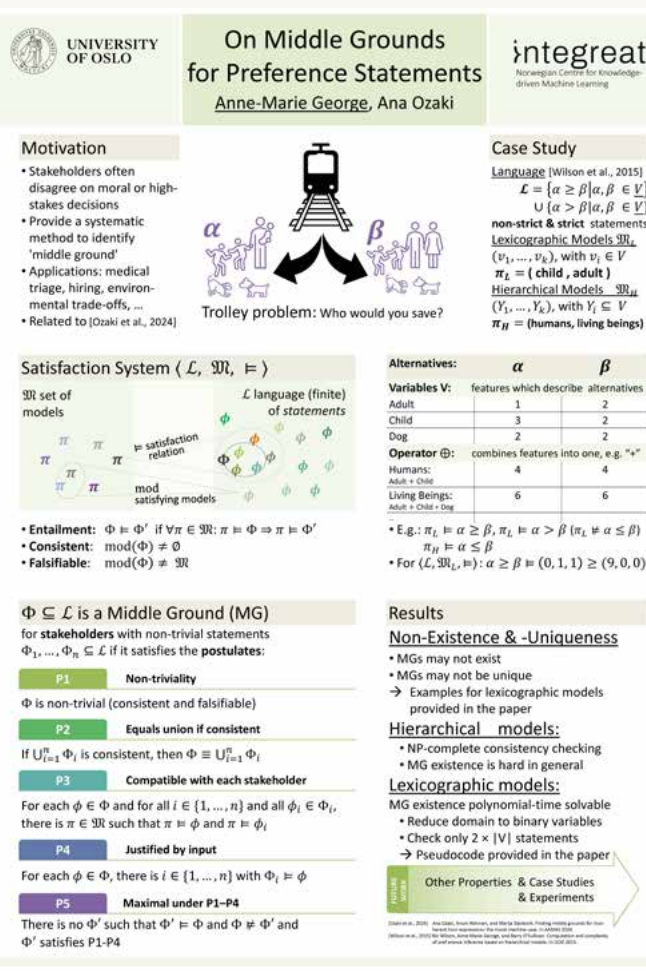
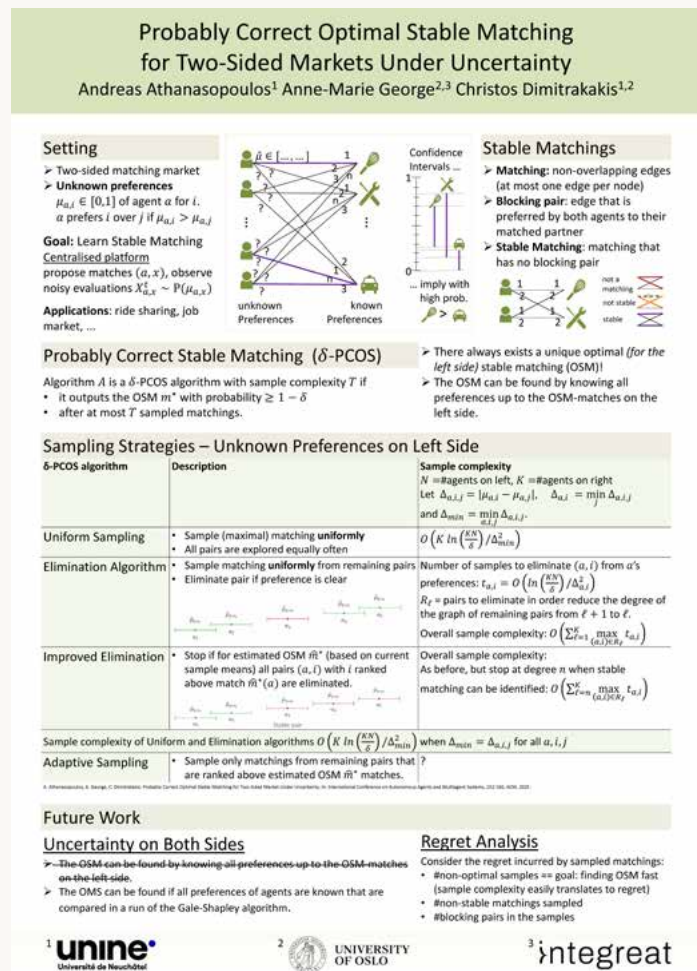
Imagine Oslo–Tromsø,
Co-writing late at night.
LLMs hallucinate
And we'll try to train them right
Imagine every meeting
Ending after light ...

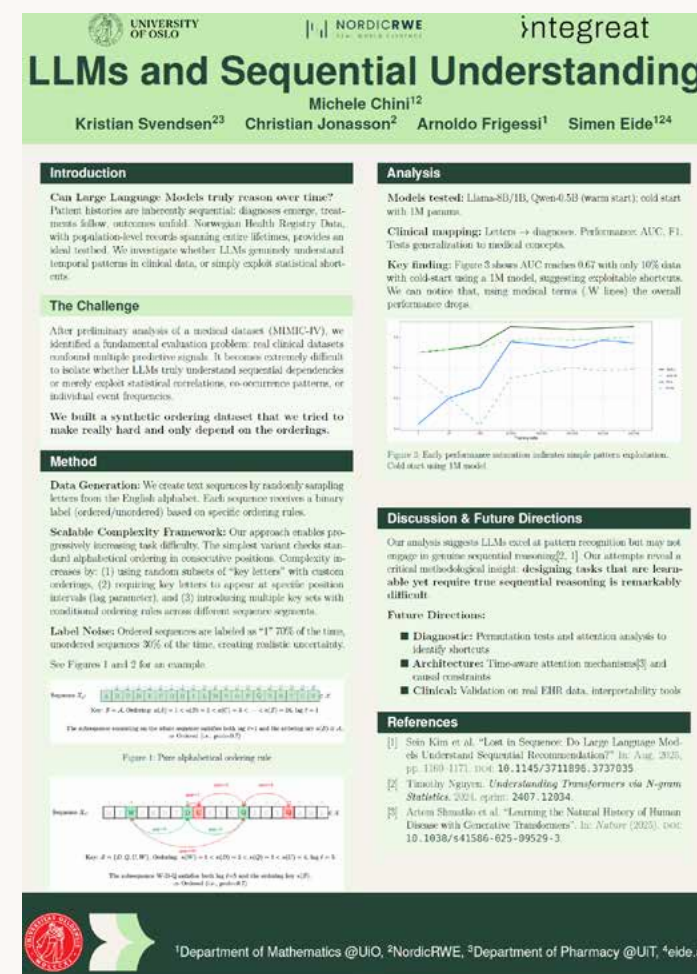
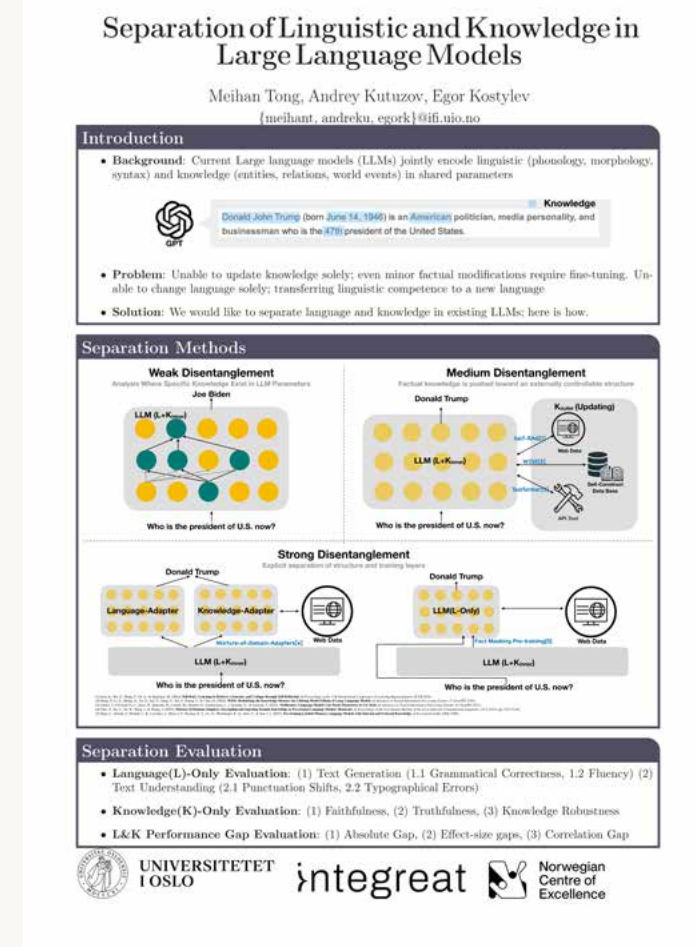
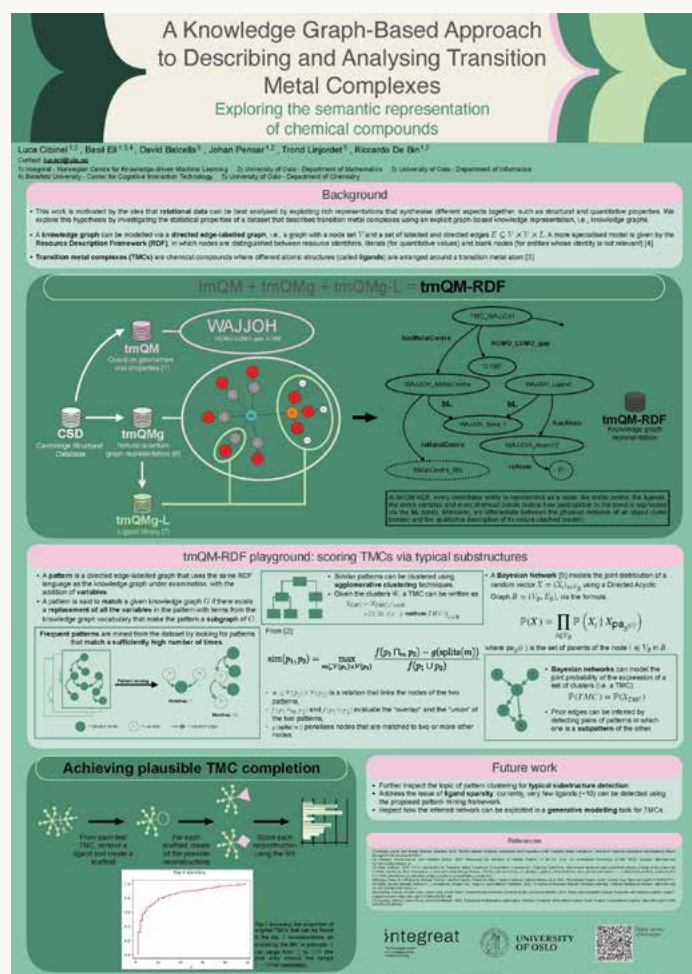
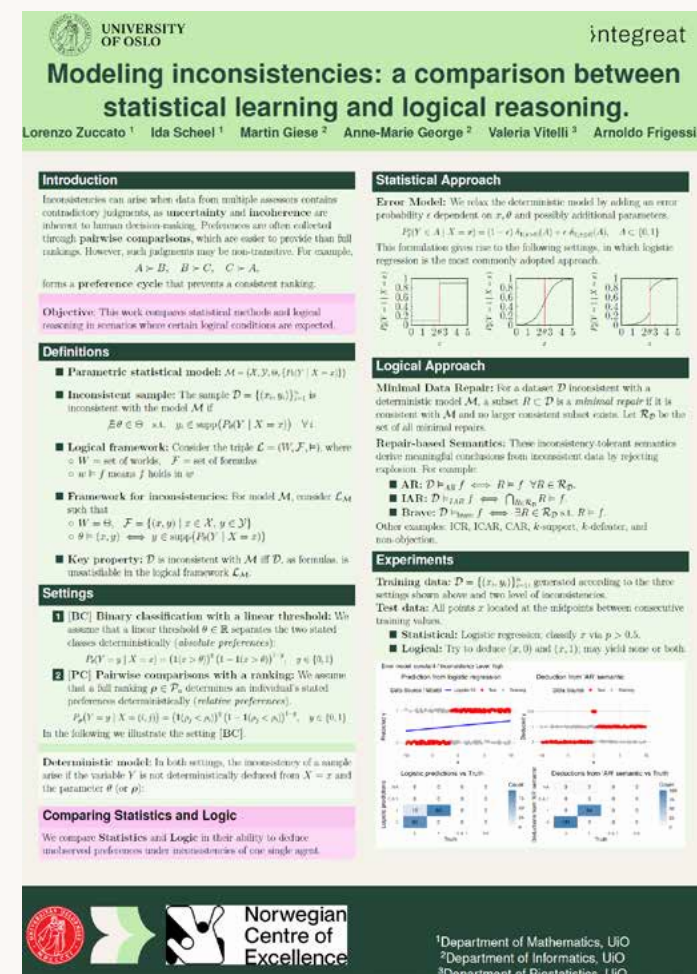
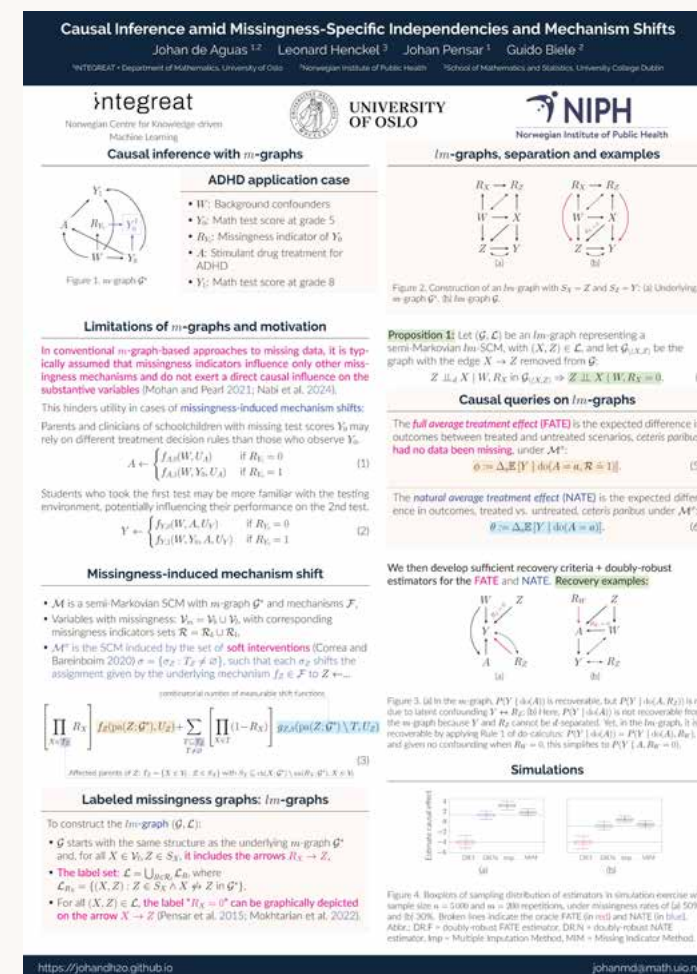
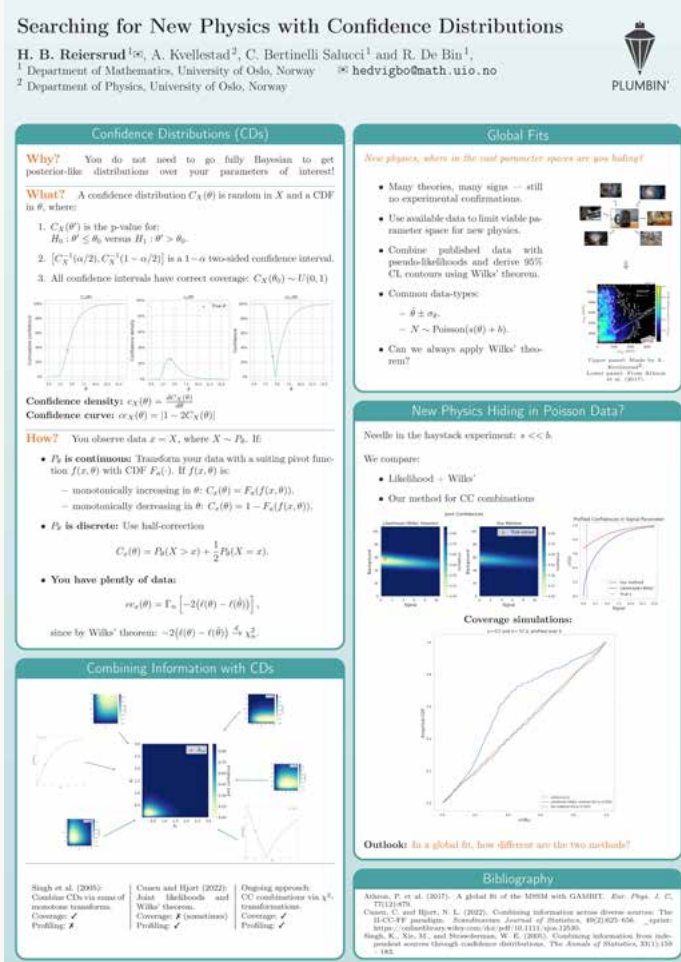
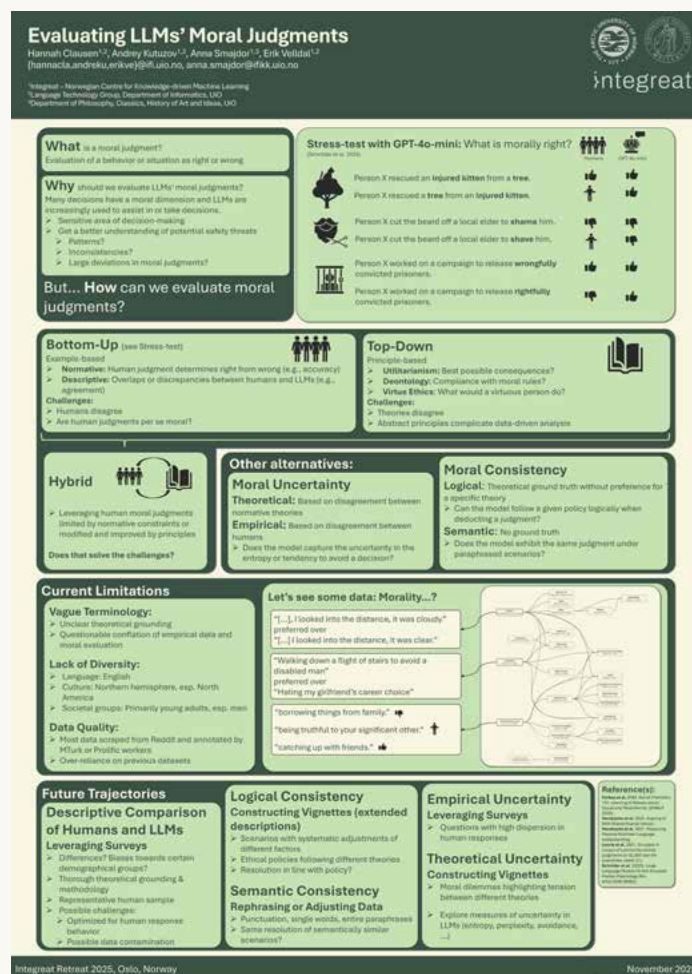
Imagine no black boxes,
Just models we can trust
No wasted computations
Algorithms fair and just
Imagine all the data,
Teaching truth, not lies ...

You may say I'm a dreamer,
But I'm not the only one.
I am happy that you joined us
And Integreat is number 1.



Selected posters presented at the Annual retreat 2025





[illegible]

TRUST in integrat

DEEPXPLAIN

Aligning attention with human rationales in hate speech detection

Brage Eilertsen, Raskva Bjergfjordstøttir, Francielle Vargas and Ali Ramezani-Keibrya

Supervised Rational Attention
jointly optimizes classification and attention-rationale
alignment.

Trained BERT-models on HateXplain and HateBRXplain
with 2.4x better explainability

Read more

USP

Uio

UNIVERSITY
OF OSLO

UIT

University of
Norway

integrat

Bootstrapped Information Bottleneck

Sigurd Holmsen^{1,2} Sylvia Richardson¹ Arnaldo Frigessi^{1,2} Ali Ramezanian-Jelani²

Abstract

The Information Bottleneck (IB) balances efficiency (sparsity, accuracy) and input compression. In this work, we use point estimates of these quantities when selecting the best model. We introduce the Bootstrapped Information Bottleneck (BIB), which introduces an ensemble of “reference” vs “compressor” autoencoders with 50% overlap. BIB allows better selection of IB models and more accurate estimates for desired levels of sparsity and compression. Compared with standard point estimates, BIB requires less computation and less data to train the ensemble.

Method: Information Bottleneck

We study representations of T inputs X using a parameter information about Y (also T inputs) Z . The Information Bottleneck (IB) objective is to

$$\max_{\theta} I(X; Z) - \beta I(Z; Y) \quad (1)$$

with a controlling parameter β (sparsity vs compression), where the L th-order Lagrangian loss

$$L_{\beta}^L(X; Z, Y) = -H(Y, Z) - \beta \cdot I(Z; Y)$$

and average gain of L (to learn to use the parameter estimate $I(Y; Z)$ to $I(X; Z)$).

Exact numerical information is intractable for high-dimensional data, so we use the Variational IB (VIB) with random q -bit and decoder p -bit.

$$\max_{\theta} I(X; Z) - \beta I(Z; Y) \quad (2)$$

Variational ML estimation on binary and high-dimensional data, offering to standardize ML estimation on ensemble of compressions.

Bootstrapped Information Bottleneck (BIB)

Originally, ML estimation on ensemble of compressions (sparsity, accuracy, compression level), we introduce the Bootstrapped Information Bottleneck (BIB, for each input T).

1) Generate bootstrap samples $S_1, \dots, S_B$ from a validation set.

2) Compute bootstrapped ML estimates

$$\hat{I}_B(X; Z) = \frac{1}{B} \sum_{b=1}^B I(S_b; Z_b) \quad (3)$$

3) Form pairs $\hat{I}_B(X; Z)$ and $\hat{I}_B(Y; Z)$ adjustable offsets

$$e_X = (x_1, \dots, x_{|S|})^T, \quad e_Y = (y_1, \dots, y_{|S|})^T$$

4) Offset: compressive loss $\hat{L}_B(X; Z) = \hat{I}_B(X; Z) - e_X$

$$\hat{L}_B(X; Z) = \hat{I}_B(X; Z) - e_X \quad (4)$$

We select T only T with 95% confidence, and BIB estimates $\hat{I}_B(X; Z)$ is instead sparse, scalable, multi-bit, ensemble ML estimates with joint uncertainty, multi-bit, multi-target, statistically guaranteed ML model selection.

References

1. Holmsen, S., Richardson, S., Frigessi, A., Ramezanian-Jelani, A. (2024). The Bootstrapped Information Bottleneck. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Press.

2. Holmsen, S., Richardson, S., Frigessi, A., Ramezanian-Jelani, A. (2024). The Bootstrapped Information Bottleneck. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Press.

Results

Figure 1: Information (in compressing the detailed model $\hat{I}_B(X; Z)$ using 50% overlap) vs. controlling parameter β (sparsity vs. compression) that parameter estimate $\hat{I}_B(Y; Z)$ to $\hat{I}_B(X; Z)$ (see eq. 4).

The information plots in Fig. 1 shows that the ML point estimates can be easily selected: the red curve shows curves to select the efficiency target, the bootstrap offsets range large, high compression, and falling below the threshold, BIB instead shows the smaller β where 95% have found either the target, large sparsity and compression levels. Table 1 shows that this leads to better generalization. In bootstrapped, selected models with our best target efficiency requirement, on the test set and also against large efficiency, according to the estimate variance (the yield) both statistically valid and better performing. BIB

Table 1: Test performance of β chosen by point estimates or BIB, Target (Frigessi et al., 2024)

	Reference	β	0.01	0.05	0.1	0.5	1.0	5.0	10.0
Sparsity	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Compression	No	Yes	No	Yes	No	Yes	No	Yes	No

Figure 2: Bootstrapped information differs across different β

Conclusion

We showed that T only T estimates in the Information Bottleneck can be made stable, leading to sub-optimal model selection. Our Bootstrapped IB (BIB) replaces point estimates with joint uncertainty results, providing statistically valid and better performing. BIB, statistically selects better models and improves downstream performance.

Norwegian
Centre of
Excellence

Integrat
Department of Informatics, UIO

Synthetic Medical Data: What is at stake?

Yael Friedman & Anna Smajdor

Introduction

With rapid advances in AI, medical applications—from diagnosis to drug development—face a growing demand for large, comprehensive datasets that are often constrained by privacy concerns, regulatory frameworks such as the GDPR, and the scarcity of cases for rare or undiagnosed conditions. Synthetic medical data (SMD) has been introduced to address data-requirements.

Three primary methodologies for SMD: Generative models to privacy guarantees, Differential privacy models, and GANs couple models. All are vulnerable (to different degrees) to inference attacks that can regenerate real SMD, SMD narratives often frame the technology as ethically superior.

Tradeoffs and Value Hierarchies

Privacy over transparency:

SMD can obscure the specific characteristics and nuances of real patient populations, hindering explanation goals that undermine clinicians' ability to interpret model outputs, erode trust, hinder bias mitigation, and risk long-term de-skilling and staled medical learning. These risks are especially acute for rare diseases, where small original datasets and overestimated synthetic approaches can lead to critical dependencies and idiosyncratic features, producing misleading clinical insights and impeding innovation and patient-centered decision-making.

Scalability over fidelity:

developers aim for low fidelity (privacy) but high distributional fidelity, yet niche and selective-distortion methods can break crucial dependencies or entrenched biases from the original data, generating a "garbage in, garbage out" problem, or even introduce new biases. This preference for scalability, sacrificing fidelity over data individual, narrated experiences, risks alienating patients, erasing contextual and minority experiences, amplifying diagnostic disparities, and institutionalizing these trade-offs as the new "normal" through a subtle "value capture."

Discussion & Conclusions

This intensified abstraction makes individuals fungible, accelerates commodification of health information, and, by erasing contextual nuances and key dependencies, undermines medical epistemology and entrenching misdiagnosis or incomplete knowledge as new normal. Datafication produces a partial, technical form of objectivity by translating meaning into decontextualized, transportable data points; SMD goes a step further, generating "zombie data" that dissolves reflection by stripping lived context and agency from those experiences.

Cell Niche Discovery via Association Rules Mining

Youssef Wally, Martin Giese, Basal El, Benjamin Roudot, Elisabeth Weisner*
*UT, The University of Nancy
University of Da

Cell Niche Discovery using Association Rules (AR) Mining

- AR describe how often items occur together
- Originate from market basket analysis
- Find relationships, correlations or patterns within large sets of data items
- Support bio-applications:
 - Cell Niches (Communities)
 - Relationship among cells

Discovery of cell niches:

Cell type 1 + cell type 2 → cell type 3 = Association Rule (AR) 1

Cell type 4 + Cell type 5 → cell type 6 = Association Rule (AR) 2

Do AR1 + AR2 have implications?

Cell type 1 + Cell type 2 + Cell type 3 + Cell type 4 + Cell type 5 → cell type 6

The Tumor Microenvironment (TME)

- Immune cells, blood vessels, extracellular cells, and molecules that interact with a tumor
- Can suppress or promote cancer depending on its composition and interactions.

Multiplexed Images

- Up to ~50 different parameters simultaneously in tissues
- Many cellular subtypes exist
- Marker cell markers unknown

Hierarchical Relationships

- Uncertainty increases at deeper hierarchy levels
- Events occur in impact
- Data-driven knowledge point

Central Questions

- How to extend association rules mining to be flexible with respect to biological relationships of cells
- Spatial and hierarchical association rules
- Building **knowledge graphs** describing the TME
- Utilizing association rules along with GNNs

Ongoing Work

- Support: measures how frequently the interaction appears
- Confidence: Probability that transactions with X also include Y
- Lift ($X \rightarrow Y$) = Confidence ($X \rightarrow Y$) / Support(Y)

Support and Lift of Association Rules

Association Rule	Support	Lift
$X \rightarrow Y$	0.1	0.5
$X \rightarrow Z$	0.2	0.4
$Y \rightarrow Z$	0.3	0.6
$X \rightarrow Y \rightarrow Z$	0.05	0.7

Support and Lift of Association Rules

Association Rule	Support	Lift
$X \rightarrow Y$	0.1	0.5
$X \rightarrow Z$	0.2	0.4
$Y \rightarrow Z$	0.3	0.6
$X \rightarrow Y \rightarrow Z$	0.05	0.7

References

- Henrik Hogeweg, Kenneth Forde, Iratjeh Wong, Lisa Allen, Anne Oelmann, and Vladimir Mikheev. Development of 42 marker genes to identify cancer-associated fibroblast niches in breast cancer using imaging mass cytometry. *Frontiers in Immunology*, 10:1207-19, 2019.
- Thomas Schaefer, Gregory D. Kulkarni, and Alessandro Lodi. Discovery, analysis, and presentation of relations in networks: Knowledge Discovery in Data Science (1998).
- Avril, John M., and Huiya Huang. Mass cytometry in hematologic malignancies: research highlights and potential clinical applications. *Frontiers in Immunology*, 11:2020, 2020.
- Fluigent Corporation.



UNIVERSITY OF AMSTERDAM
Vrije Universiteit Amsterdam

Flatland and Beyond: Mutual Information Across Geometries

Youssef Wahby¹, Johan Mylius-Kuusik², Michael Kampferer³, Rezaan Ehsani⁴, Vladen Mislove⁵,
 Elisabeth Wehrer⁶
¹Department of Physics and Technology, ²UT The Arctic University of Norway
³Norwegian Computing Center
⁴Department of Clinical Medicine, University of Bergen



Flatland

Image Microscopy



Cell Segmentation



Cell Classification



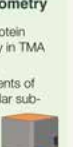
Neighborhood Classification



- Tumor microenvironments in breast tissue are captured by imaging mass cytometry
- Cells are segmented based on nuclear markers
- Cells are classified hierarchically in approx. 50 classes
- Neighborhood Descriptors can be defined in several ways to identify non-random spatial associations

Imaging Mass Cytometry

- Up to ~50 different protein targets simultaneously in TMA cores
- 10 breast cancer patients of three different molecular subtypes
- High spatial context of the tissue
- 20 subsets of each immune cells and fibroblasts have been identified

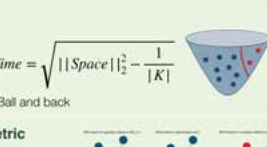


Lorentz Space

- Dimension: $\{X_0, X_1, X_2\}$
- For a K-constant space: $Time = \sqrt{||Space||^2 - \frac{1}{K}}$
- Easily project to Poincaré Ball and back

Geometric Affinity Metric

- $D(x, y) = \frac{1}{\sqrt{|E|}} \arccos\left(1 - \frac{|K| < x, y > |}{|E|}\right)$
- $x, y > y = (Time_x - Time_y) + (Space_x - Space_y)$
- $f_{\text{geo}}(x, y) = \phi(x) + \phi(y) - \frac{1}{N} \sum_{i=1}^N \phi(x_i) - \phi(y_i) \cdot D(x, y)$



Study Outline

- We introduce an information-theoretic framework based on nearest neighbor estimators to rigorously quantify the clustering performance in each geometry using mutual information and conditional mutual information.

Quantity

$MID(x, C)$

$CMID(x, C)$

$CMID(x, D)$

$CMID(x, C | D)$

IMC - IMC

IMC

IMC

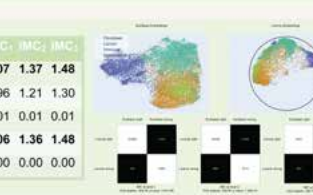
IMC

IMC - IMC

IMC

IMC

IMC



References

1. Wang, Y., Yu, T., Mouton, A., Shi, H., & Zhou, G. (2021). Hyperbolic deep neural networks for cell classification. In Proceedings of the AAAI Conference on Artificial Intelligence (pp. 14212-14219).

2. Bollen, A.M., Winter, S., Scherf, A., & Neis, J. (2019). A multi-scale approach to cell classification: from single cells to tissue. In Proceedings of the AAAI Conference on Artificial Intelligence (pp. 14212-14219).

3. Koushki, A., Arora, H., Bhatnagar, S., & Poon, H. (2019). Hyperbolic deep neural networks for cell classification. In Proceedings of the AAAI Conference on Artificial Intelligence (pp. 14212-14219).

4. Wang, Y., Yu, T., Mouton, A., Shi, H., & Zhou, G. (2021). Hyperbolic deep neural networks for cell classification. In Proceedings of the AAAI Conference on Artificial Intelligence (pp. 14212-14219).

5. Mislove, V., & Ehsani, R. (2021). Development of 42 cancer panels for deep study of cancer associated biological traits in breast cancer using multi-scale spatial analysis. Frontiers in Immunology, 12, 625741.

paper

githu



Flatland

Launching TRUST Integreat in a Growing AI Landscape

In 2025, the new AI centre TRUST – The Norwegian Centre for Trustworthy AI was established, with Arnaldo Frigessi serving as Director while continuing as Co-director of Integreat. The establishment of TRUST marks an important development in Norway's AI landscape and situates Integreat within an expanding national research environment on trustworthy AI.

For Integreat, this development reinforces the relevance of our long-term focus on knowledge-driven machine learning.





Questions of trust, robustness, and responsibility have been integral to the centre's research from the outset. The emergence of complementary initiatives such as TRUST underscores the importance of these themes and strengthens the broader context in which our work is carried out.

At the same time, Integreat maintains its distinct profile: integrating general and domain-specific knowledge into machine learning through interdisciplinary collaboration. The continued involvement of shared leadership ensures coordination and dialogue, while allowing the centres to pursue their respective priorities.

Rather than representing a shift in direction, this development confirms the trajectory Integreat has already taken and opens additional opportunities for collaboration and exchange.



TRUST – The Norwegian Centre for Trustworthy AI is an AI centre funded by the Research Council of Norway (2025–2030) and part of the six national AI research centres. It is dedicated to developing artificial intelligence that is accurate, interpretable, inclusive, fair, safe, sustainable, and well-governed. It is hosted by the University of Oslo, Norwegian Computing Center (NR) and SINTEF, together with more than 70 partners across academia, industry, the public sector and civil society.

Integreat meets the public

Several researchers engaged in the public debate on AI and the role of science in society. Contributions in national media addressed questions of trustworthy AI, responsible research practices, and the importance of building shared infrastructure and open research traditions in the Nordic context.

Researchers also participated in discussions concerning academic freedom and the political conditions shaping scientific work. These engagements reflect Integreat's commitment to taking its societal role actively.

Integreat in the media 2025

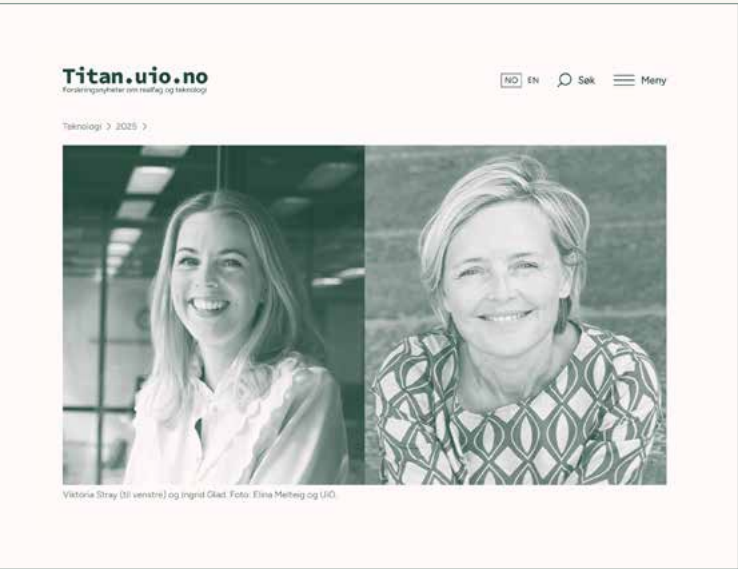


Announcement:

Ingrid Glad får millionstøtte til videreutvikling av statistiske metoder

<https://www.titan.uio.no/teknologi/2025/ingrid-glad-far-million-stotte-til-videreutvikling-.html>

(Norwegian)

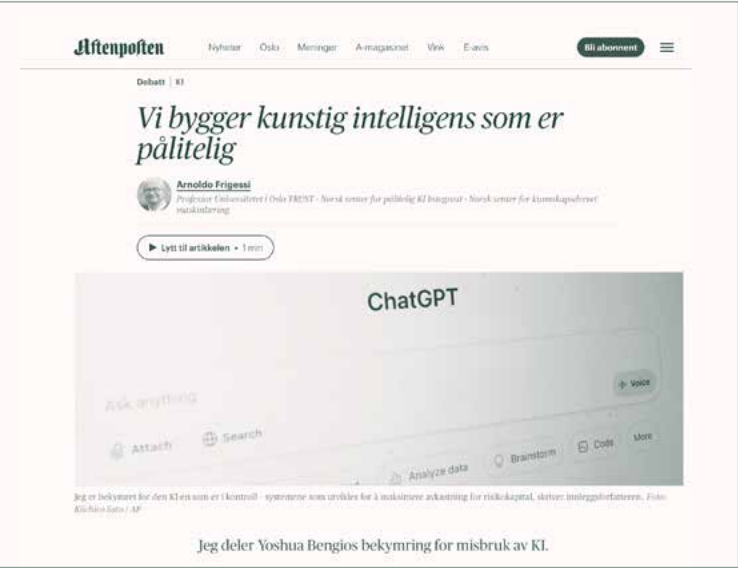


Announcement:

Viktoria Stray og Ingrid Glad er kåret til to av Norges fremste teknologikvinner

<https://www.titan.uio.no/teknologi/2025/viktoria-stray-og-ingrid-glad-er-karet-til-to-av-n.html>

(Norwegian)



Debate:

Vi bygger kunstig intelligens som er pålitelig

<https://www.aftenposten.no/meninger/debatt/i/Ok-b2ew/vi-bygger-kunstig-intelligens-som-er-paalitelig>

(Norwegian)



Glem KI-kappløpet – la oss gjøre det på vår måte!

Norden kan svare på de enorme kapitalbeløpene som investeres i amerikanske KI-giganter, investeringer som krever raske resultater, ved å bygge delt infrastruktur og satse på åpen forskning.

<https://www.dn.no/innlegg/kunstig-intelligens/teknologi/forskning/glem-ki-kappløpet-la-oss-gjore-det-pa-var-mate/2-1-1848044>

(Norwegian)



Forskere i Bergen deltok i amerikansk protest

<https://www.khrono.no/notice/949675>

(Norwegian)



Slik står du imot Trumps angrep på vitenskapen

<https://www.khrono.no/slik-star-du-imot-trumps-angrep-pa-vitenskapen/950115>

(Norwegian)



Integreat Digitalisation Lecture 2025

The Digitalisation Lecture has become a defining annual event for Integreat and the University of Oslo.



[Visit webpage](#)

Each year, the Minister for Digitalisation and Public Governance is invited to present the government's assessment of Norway's digital development and future priorities. The lecture serves as a meeting point between policy, research, and education.

In September 2025, Minister Karianne Tung addressed the question of how digitalisation and artificial intelligence can help build and safeguard Norway, in light of the national ambition to become the world's most digitalised country by 2030.

The minister highlighted three structural drivers shaping Norway's digital future: demographic change and the need for technological support in welfare services; increasing security challenges in the form of cyberattacks and digital fraud; and rapid technological development, with artificial intelligence playing a central role in value creation and modernisation.

She outlined key governmental priorities, including the development of national AI infrastructure, support for Norwegian language models, implementation of the EU AI Act in Norwegian law, strengthened regulation of data centres, and substantial investment in research and competence building. As she emphasised, digitalisation



is not merely a technical undertaking, but a societal project that must be rooted in trust and openness.

The lecture was opened by newly elected UiO Rector Ragnhild Hennum, who underscored the importance of sustained dialogue between decision-makers, researchers, and students. Integreat Director Arnoldo Frigessi highlighted the responsibility carried by those who design and develop AI systems, noting that researchers and students are central actors in shaping technology within democratic and knowledge-based frameworks.

The lecture was followed by a panel discussion moderated by Co-director of Integreat Professor Ingrid Glad, with participation from the minister and the UiO researchers

Erik Velldal and Leiv Rønneberg. The discussion addressed open-source practices, the development of Norwegian and Sámi language models, educational responsibility, and the role of trust in digital transformation.

Through the Digitalisation Lecture, Integreat provides an arena for structured dialogue between political leadership and the research community. The event reflects the centre's commitment to engaging constructively in national discussions on digitalisation, while maintaining scientific independence and critical reflection.





Integreat at Arendalsuka 2025

Integreat at Arendalsuka 2025 was focusing on the next generations. In August 2025, Integreat participated in Arendalsuka with a fully attended event in NORA's Artificial Intelligence tent. The session, titled *What's Happening with Artificial Intelligence? – The Changing Fundamental Ideas*, brought together eight early-career researchers affiliated with Integreat, who delivered short, focused presentations on emerging directions in artificial intelligence.

In four-minute talks, they addressed topics such as reasoning models, causal machine learning, uncertainty quantification, bias in algorithms, synthetic data, and the role of AI in future science. The format required clarity and precision, and demonstrated the ability of young researchers to communicate across audiences.

The event concluded with a panel discussion involving youth politicians from across the political spectrum. The discussion explored how research communities and politicians of the young generations can contribute responsibly to policy and societal development. The exchange underscored the importance of dialogue between scientific and political perspectives.

By participating in Arendalsuka, Integreat contributed to a broader public conversation about trustworthy AI, scientific responsibility, and the future of technology in society. The event reflected the centre's commitment to combining methodological depth with openness and engagement beyond academia.



[Visit webpage](#)



In four minutes each, the PhD candidates and postdocs presented:

Reasoning Models: AI that Thinks Before Acting

Håkon Austlid Taskén

Why is AI Lagging Behind in the Physical World?

August Arnstad and Sigurd Holmsen

Causal Machine Learning: Tackling Discriminatory Algorithms

Camilla Lingjærde

Uncertain Models in an Uncertain World

Hedvig Borgen Reiersrud

Why Teach AI Models to Forget?

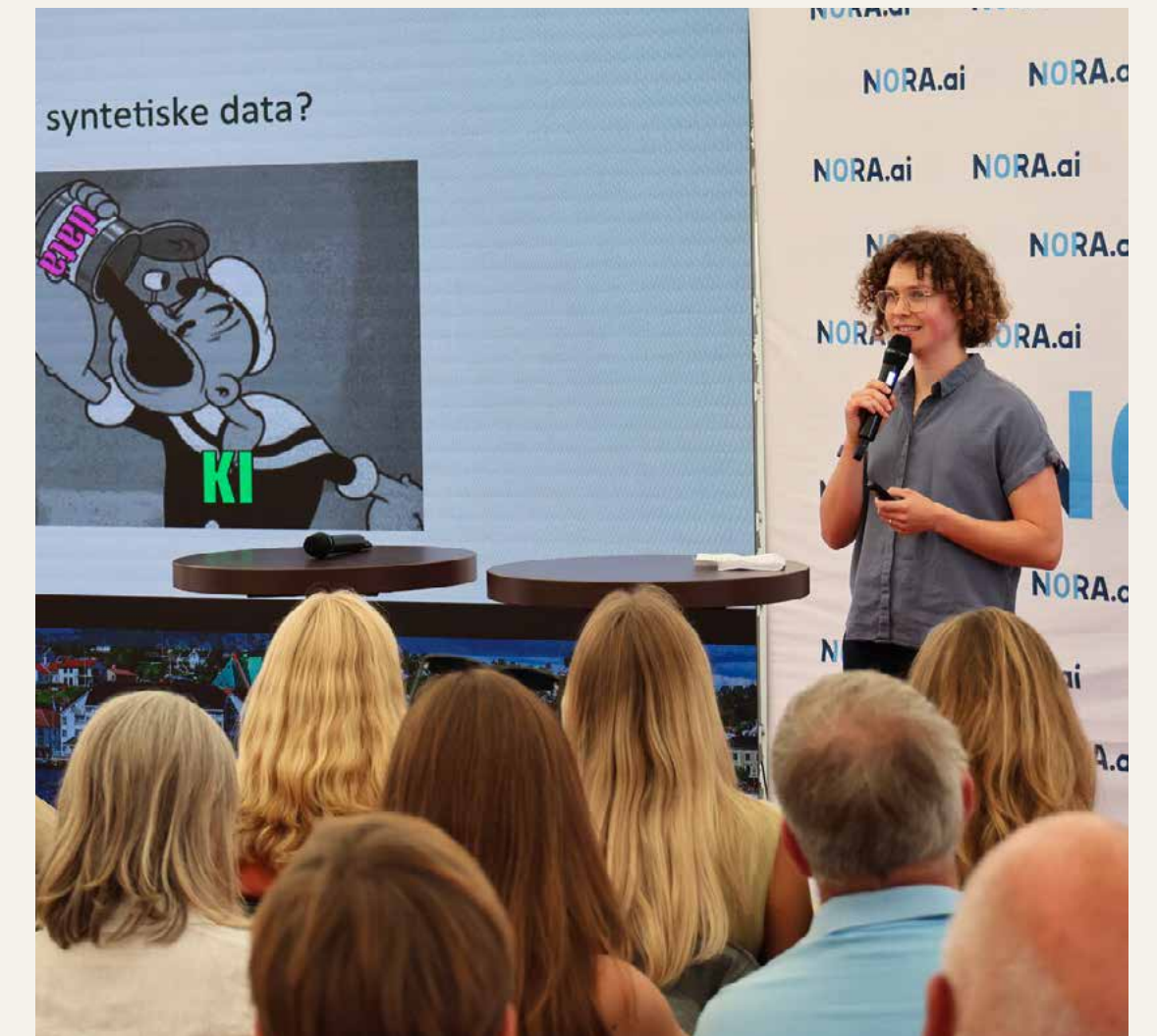
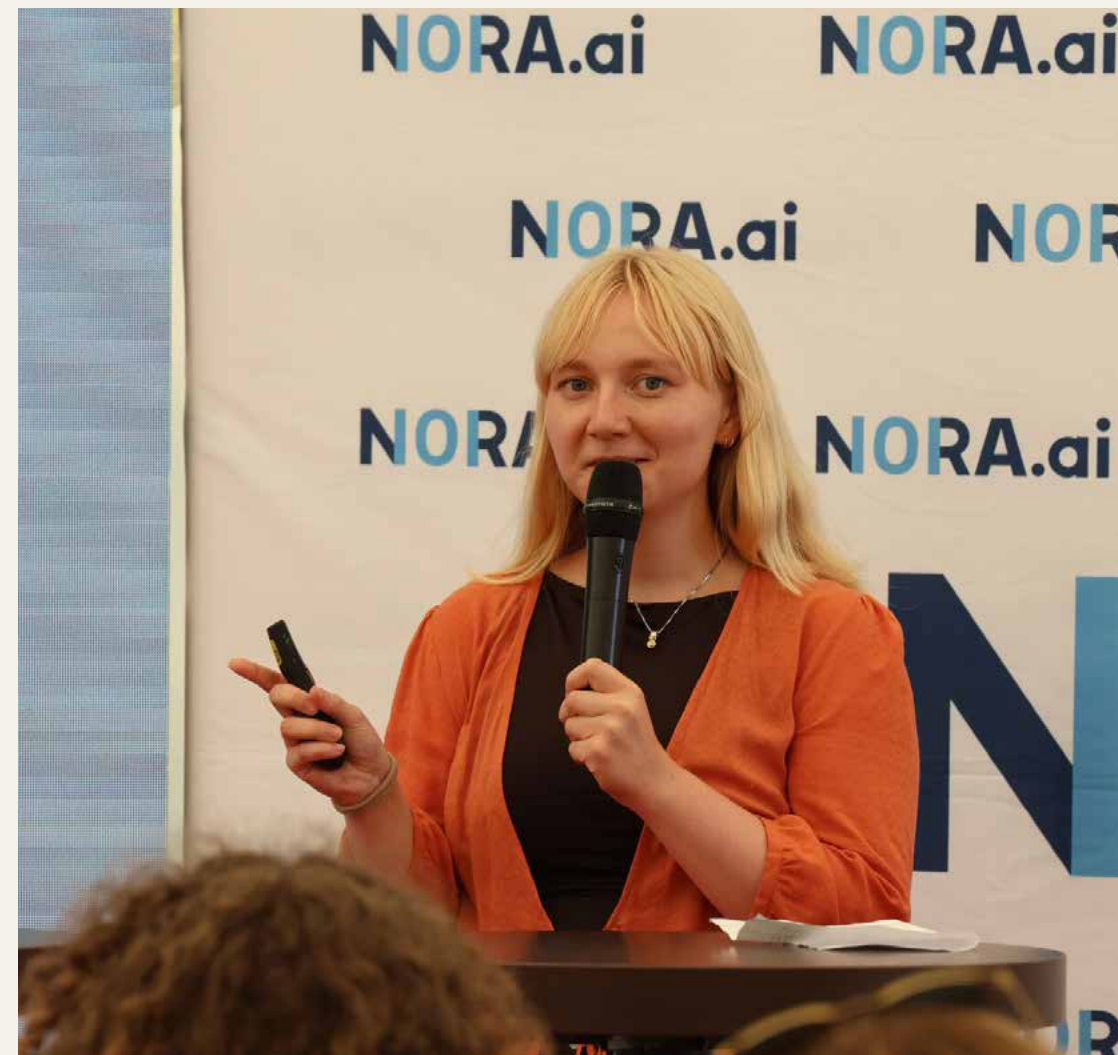
Emil Poiesz

Beyond the Real: The Role of Synthetic Data in AI Development

Elisabeth Griesbauer

Artificial Intelligence in the Science of the Future

Vilde Hansteen Ung





Researcher training in Integreat

The community of PhD candidates and postdoctoral fellows has grown to approximately 35 across the two hubs, and what began as a newly formed cohort has developed into a connected and highly productive research environment.

Those who joined in 2024 are now deeply engaged in their projects. Some fellows are beginning to plan longer research stays abroad, and the centre actively encourages international mobility as part of academic development.

Supervision remains central to the programme. As interdisciplinary projects evolve, we continue to reflect on how supervision can best support both scientific depth and cross-theme collaboration. The emphasis is on the results as well as on developing independent researchers capable of navigating complexity and contributing across disciplines.

The “research kitchens” — small, pop-up workshops designed to bring researchers together around emerging ideas — are still

developing as a format gaining clarity in the opportunities they can offer.

Academic growth of our students is supported through university courses and tailored sessions organised within the centre. These include workshops on presentation skills, strategic digital presence, and scientific communication across audiences.

Career development has become more explicitly diversified. Recognising that fellows may pursue academic, public sector, or industry pathways, the centre offers CV workshops tailored to different career directions, mock interviews, and individual career guidance provided by the administrative leadership. The aim is to equip researchers to make informed choices.

INTREF: In their own words

The INTegreat Young REsearchers Forum (INTREF) is a network created for early-career researchers within Integreat. Currently, the PhDs Roberto (Tromsø), Johan (Tromsø), Emil (Oslo), Kush (Oslo), and postdoc Clara (Oslo) volunteer as representatives from the Integreat hubs in Tromsø and Oslo. INTREF exists to create deeper professional and social connections, foster collaboration, and strengthen our research community from within, while giving young researchers a collective voice in shaping our academic environment.

Our work is structured around three main pillars: scientific development, transferable skills, and social skills. We are an open and participatory forum that welcomes new ideas and initiatives, aiming to support them from the ground up. We organize monthly assemblies on the first Wednesday of each month, where we discuss ongoing initiatives, plan activities, and address issues relevant to early-career researchers.

Initiatives arranged in 2025 by and for INTREF/Integreat researchers included a two-day research retreat to Gressholmen, where participants presented their projects and explored thematic and methodological connections to foster internal collaboration; a weekly volleyball gathering at the gym to strengthen informal bonds; a CV workshop focused on career development both inside and outside academia; a themed party each semester, open to the whole Integreat community; and an INTREF Day, ahead of the annual Integreat retreat, consisting of career-focused activities dedicated to young researchers. Looking ahead, we aim to facilitate more research kitchens and research visits, further bridge the social and academic gap between Oslo and Tromsø, support multidisciplinary collaboration, and establish stronger connections with industry through seminars and presentations.





Inclusive excellence:

a commitment to equity, diversity,
inclusion, and belonging

Our aim at Integreat is to treat inclusiveness as a natural part of our everyday work life. To produce best research results, we would like our researchers to feel safe and included. When researchers feel like equal members of a community, they are more willing to share unfinished ideas, challenge one another constructively, and explore new directions. This openness shapes how creative we can be together and how confidently we push the boundaries of our field.

Inclusive excellence

I am very grateful that at Integreat equality, diversity and inclusion are held high in times where these values are up for discussion or openly under attack. This fall's DEI's focus was on mental health in academia. Even though healthy minds are at the foundation of good academic work and many of us will at some point in our lives be directly or indirectly affected by struggles with mental health, mental health is still too little talked about openly. The resonance to creating a space for sharing personal experiences, vulnerabilities and reflections in our mental health seminar was accordingly great and positive. Becoming aware that this is a shared experience and feeling the support of not just individual colleagues but the Integreat community, made and makes us a more compassionate and resilient community. — Integreat Inclusion Officer PhD Elisabeth Griesbauer

Research is often perceived as a solitary activity, but much of our best thinking emerges in conversation, in shared ideas, and around the lunch tables in our lounge. We therefore value the everyday interactions, openness, and sense of belonging that allow ideas and people to grow and dare.

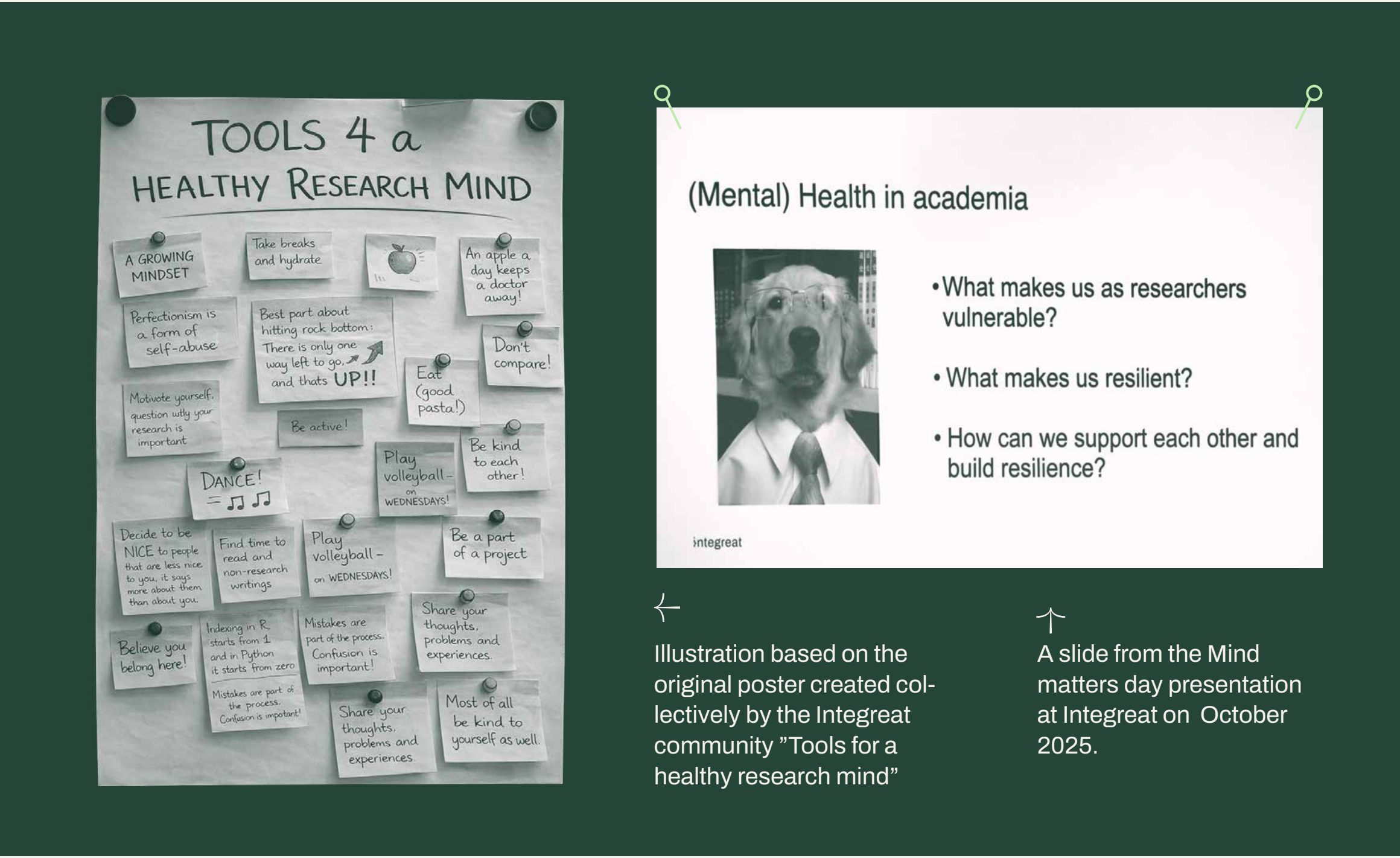
In a research environment like at Integreat, shaped by interdisciplinarity and rapid development of ideas, differences in background, discipline, and perspective are a strength. Therefore, scientific excellence depends on our ability to bring these perspectives together in ways that allow everyone to contribute meaningfully.

In 2025, this commitment to inclusiveness has taken shape in several ways and throughout the whole year.

We monitor our recruitment processes closely. For instance, approximately 30% of our early-career

researchers in the Intref group are now women, reflecting gradual progress. However, we are not yet where we want to be. In 2026, we will have more emphasis on our efforts to extend beyond recruitment. We are increasingly focused on understanding what shapes people's experiences after they arrive — what supports them, what challenges them, and what influences whether they choose to remain and develop within the centre. This includes attention to supervision, career pathways, and the everyday working culture.

We have strong female role models in both research and leadership. These researchers actively support women in science and contribute to initiatives that encourage girls to pursue STEM. At the same time, our guiding principle is that opportunities and conditions should be fair for everyone, regardless of gender or background. Representation and structural fairness matters alike.



← Illustration based on the original poster created collectively by the Integreat community "Tools for a healthy research mind"

↑ A slide from the Mind matters day presentation at Integreat on October 2025.

The impediment to action advances action. What stands in the way becomes the way” (from Marcus Aurelius in the Meditations). I think this says a lot about what we actually do in our everyday life of researchers, when we struggle and struggle and struggle and we kind of get frustrated and think we will never see the light; but the reality is that in that unsucces and frustration we are already building the roots of the solution, just we will only know it in the end. And only a posteriori we will be able to recognize that the struggle was necessary. In Integreat, this resonates with the challenges and slowness we face every day when trying very hard to communicate to each other across diverse cultures, for example in interdisciplinary projects. The first two years of Integreat we struggled a lot, but now I see that this effort was really fruitful. — PI Valeria Vitelli

Inclusiveness reflects in our everyday culture. As part of the World Mental Health Day, we set aside time to reflect on resilience and the realities of research life. Academic work also involves uncertainty, setbacks, and high expectations. We believe that creating space for reflection and shared understanding is one way of acknowledging that excellence and well-being are connected. The session, led by Integreat’s Inclusion Officer, PhD Fellow Elisabeth M. Griesbauer, focused on resilience, openness, and navigating the fluctuations inherent in academic work. The event created space for shared reflection and collective awareness. The initiative reflects the centre’s understanding that mental resilience is not an individual trait alone, but a property of research environments that allow space for dialogue and mutual support.

Finally, questions of inclusion are present in the research itself. Workshops on bias and inclusive research cultures, together with projects examining social bias in large language models, remind us that fairness is not only an organisational issue — it is also a scientific one. How we build models, how we evaluate them, and whose perspectives are represented are all part of responsible research practice.

“For us, inclusiveness is an ongoing responsibility — something we return to as we recruit, supervise, collaborate, and develop knowledge together”.





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.

Integreat in conversation with practice

Integreat researcher Leiv Rønneberg held a three-day course on the mathematics behind machine learning (ML) aimed at upper secondary maths teachers in Oslo.

About the course

The course was a collaboration between Integreat, the unit for Research, Innovation and Competence Development in Schools (FIKS), and Oslo municipality as part of FIKS and Oslo municipalities existing schools partnership. The idea for the course came from Knut Skrindo, a teacher at Oslo cathedral school, who at the end of each day also led discussions on how and if ML can be anchored within existing maths curriculum.

The course ran over three afternoons, starting from basic statistical models such as linear and logistic regression, and adding more complexity until the group

arrived at the single-hidden layer neural network - which was studied in detail on day 1. On day 2, deeper networks were motivated by the need to fit more flexible functions and how that can be achieved by deeper nets such as the multilayer perceptron. The backpropagation and stochastic gradient descent algorithms were introduced before moving on to more complex neural networks such as recurrent and convolutional architectures. On the final day the focus shifted towards the attention mechanism and transformer architecture which drives modern large language models, as well as denoising diffusion models used in image generation.

The emphasis in the course was on the underlying mathematics of machine learning, and understanding in detail the small building blocks that make up modern machine learning architecture, rather than understanding the large models themselves. The teachers were also provided with example code for training their own neural network models, and the course drew around 25 participants from various schools around Oslo.



I learned an incredible amount about AI. It was rewarding to engage with material that was challenging, yet still felt manageable. I feel I have genuinely gained further education and become more intellectually enriched.

— one of the math teachers at the course

Impact

For the teachers, the course provided conceptual tools to approach machine learning not as a black-box technology, but as mathematics that can be unpacked, questioned, and connected to existing curriculum. By working through regression models, neural networks, and optimisation algorithms step by step, participants were able to situate modern ML within topics they already teach — functions, derivatives, linear algebra, and probability. This supports more confident and critical classroom discussions about AI and its mathematical foundations.

For Integreat, the course illustrates how research expertise can translate into competence development without simplifying the subject matter. It also strengthens long-term collaboration with schools and contributes to building mathematical literacy around AI at the upper secondary level. In this way, outreach is not separate from research, but an extension of it into educational practice.

It is sometimes said that teachers are also students. Our experience at Integreat suggests that this is true.



AI for Justice: Using Generative Artificial Intelligence (AI) in War Crimes Archiving

The workshop was conducted in a ‘Hackathon’ format by the UiT-hub of Integreat in June 2025. It involved various tutorials on how to prompt efficiently, use deep research tools and code with AI.

I’m really happy with the workshop and grateful to the organisers for putting together such a well-structured and insightful event. The knowledge and skills I gained will definitely be useful for the work of my organisation, and I’m looking forward to applying them in practice. It was also a pleasure to meet and exchange ideas – such a productive and inspiring experience! I’m already looking forward to seeing all of you at the next workshop.

— a workshop participant

The workshop brought together a broad range of representatives from academia, civil society, and expert communities interested in working with online and offline war crime databases and archives. Over the two-day hackathon, participants familiarised themselves with recent Generative AI tools and learned how to apply them by completing practical assignments using various digital materials. Guided by the Integreat researchers and through individual tasks and teamwork in mixed interdisciplinary groups, participants shared their experiences and built a small ‘AI-aware community’, which will further

specialise in using Generative AI and continue training together to advance effective systematisation, preservation, and presentation of war crime archives.

Due to the proliferation of international conflicts and the accumulation of large volumes of evidence on potential crimes, it is crucial that representatives of academia, the public sector and civil society work together to develop common standards necessary to systematise the collected evidence into accessible and searchable digital archives.

I was astonished by the enthusiasm of the participants toward AI, and their high motivation to learn. There is definitely high interest in AI, and I am sure it will have a strong positive impact for justice.

— Integreat co-director Benjamin Ricaud

While participants learned new AI tools for archiving war crimes and created new professional networks, the organisers received an overview of challenges in the field and established a practitioners' community interested in using Generative AI. Due to the identified challenges and the networks built as

part of this workshop and together with an international consortium of partners, the organisers developed a project proposal under the Horizon Europe Work Programme (Cluster 3: Civil Security for Society, Research and Innovation Action) and submitted it in November 2025.



Integreat Insights: Research-Based Perspectives on AI

In 2025, Integreat launched Integreat Insights, a blog written by researchers across the centre. The initiative creates a structured space where core ideas from ongoing research are communicated to a broader audience — including teachers, policymakers, students, and others seeking a deeper understanding of artificial intelligence.

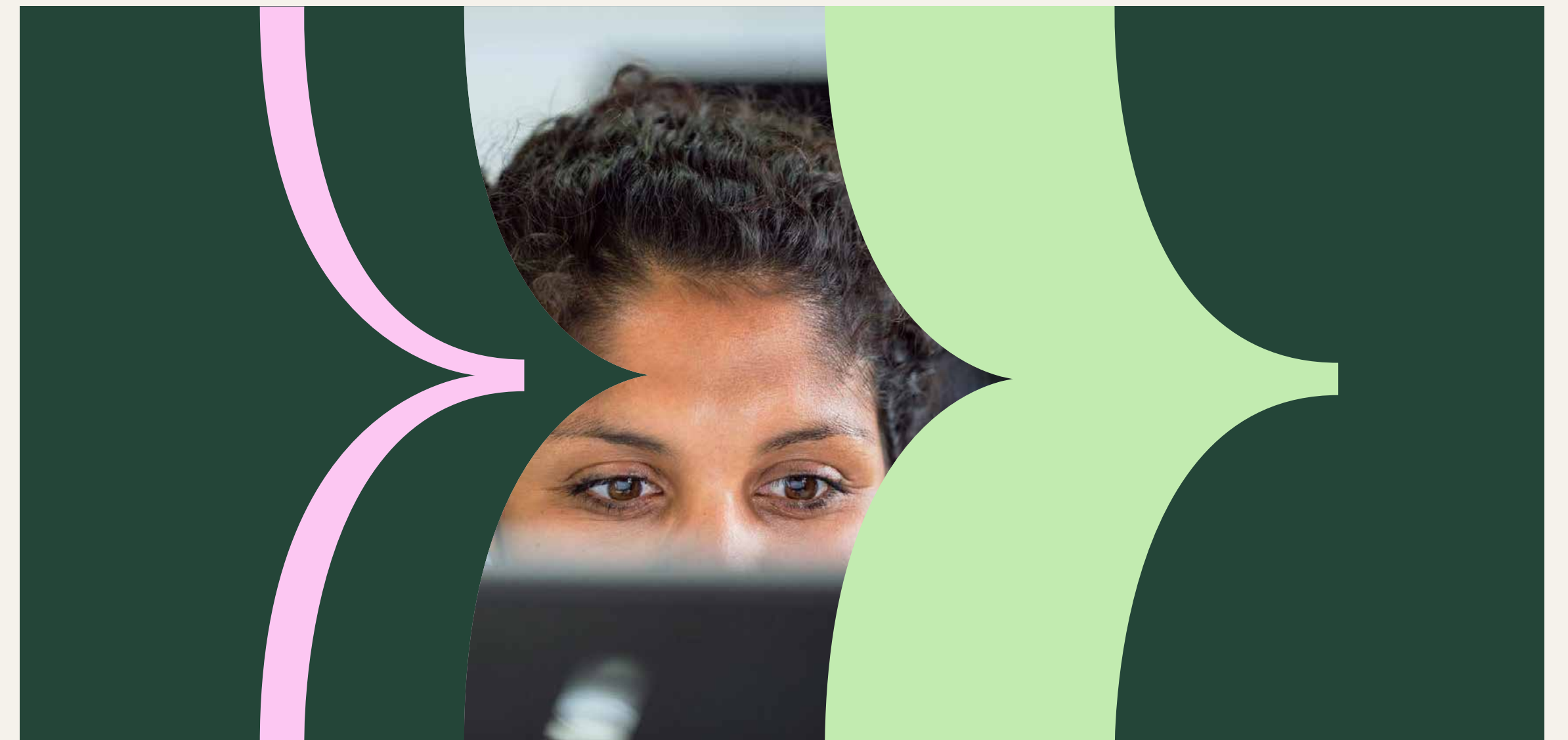
The purpose of the blog is to clarify concepts, methods, and assumptions that shape contemporary machine learning. Contributions address topics such as synthetic data, epistemological questions in AI, and the limits of large language models. By grounding public-facing essays in active research, the blog contributes research-based perspectives to ongoing debates about AI's capabilities, risks, and societal implications.

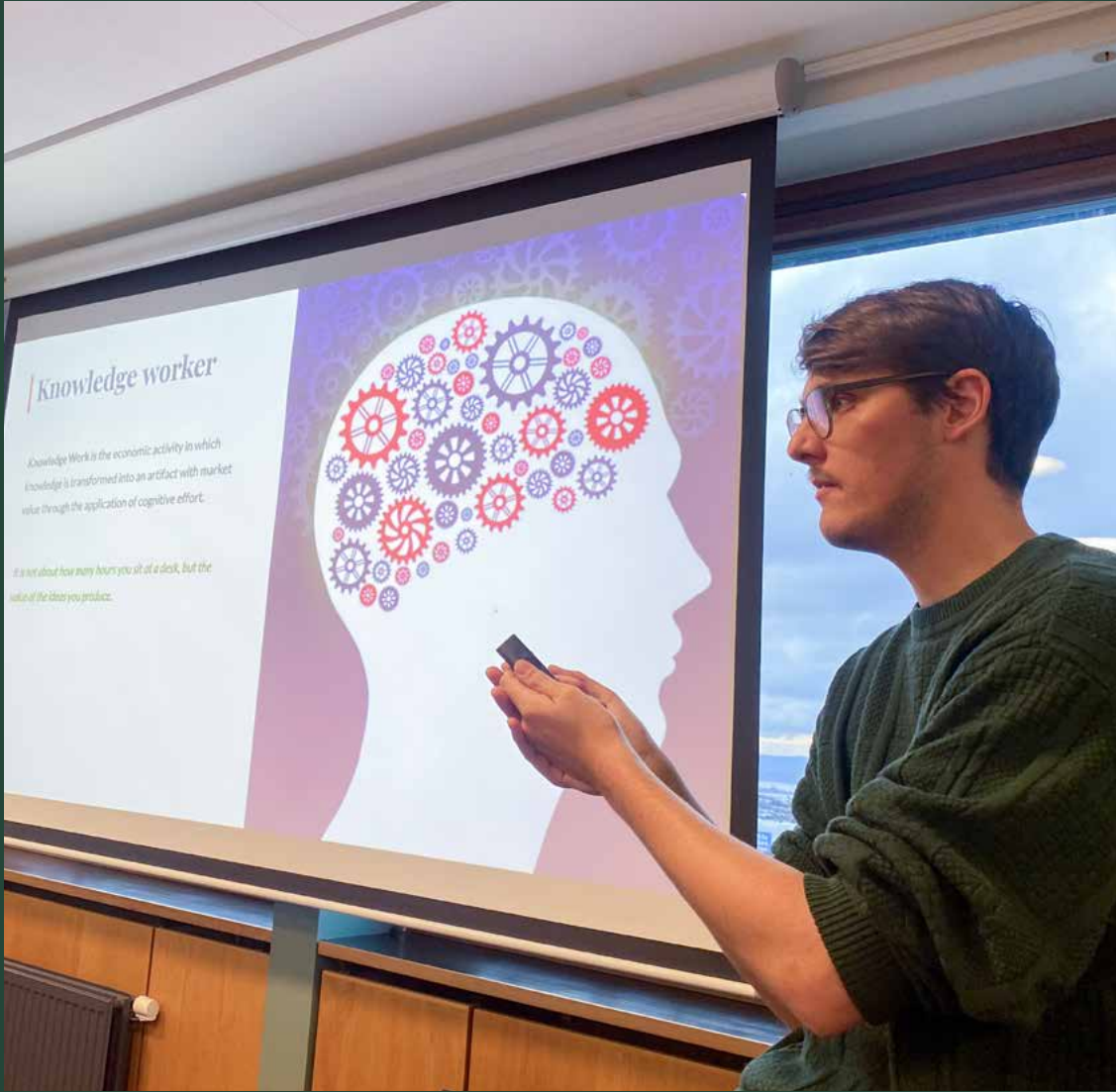
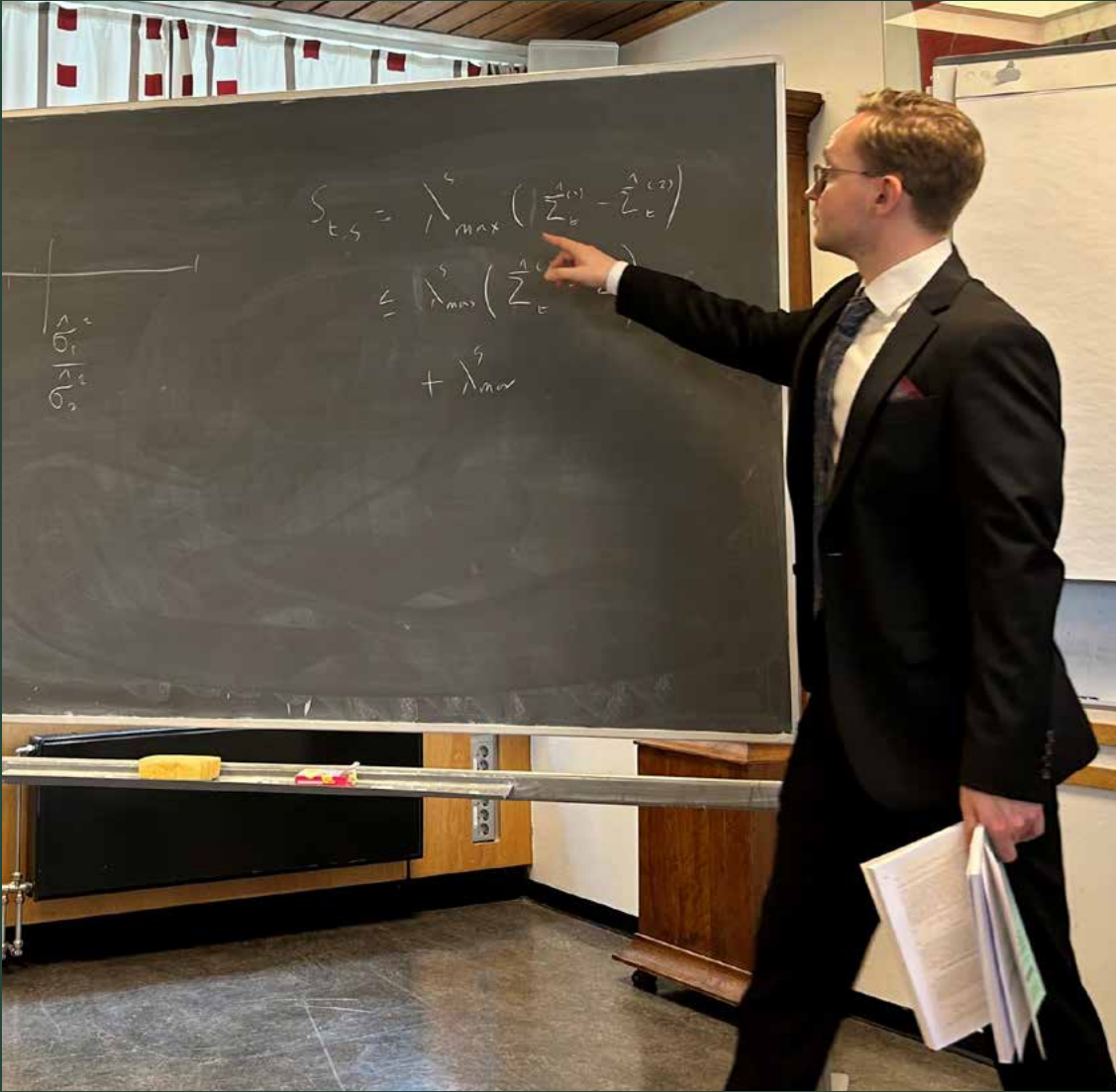
At the same time, Integreat Insights strengthens the centre internally. It provides early-career and senior researchers with an opportunity to articulate their work beyond disciplinary boundaries, making implicit methodological choices explicit and situating technical research within broader ethical and epistemological contexts. In this way, the blog extends the centre's commitment to interdisciplinarity into public engagement.

The blog forms part of Integreat's broader effort to foster informed and responsible discussions about AI, grounded in methodological rigor and conceptual clarity.



[Visit webpage](#)







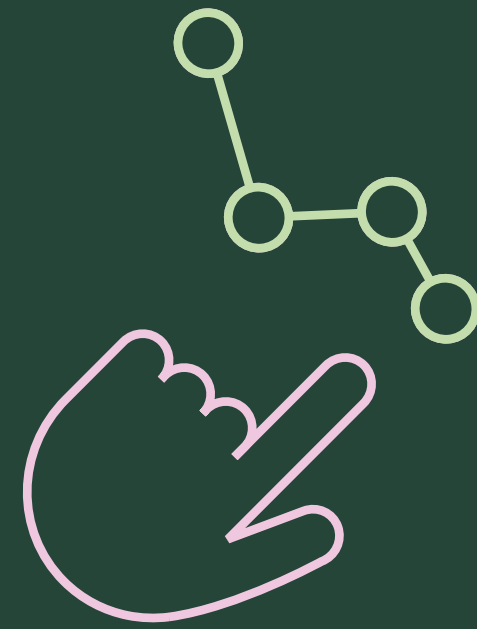
Integreat in numbers

Research Council of Norway funds:

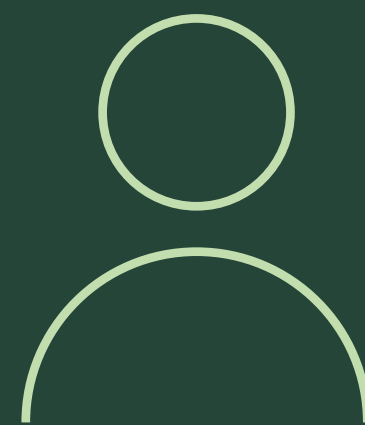
17 255 950

Own contribution and cash partners:

20 769 157



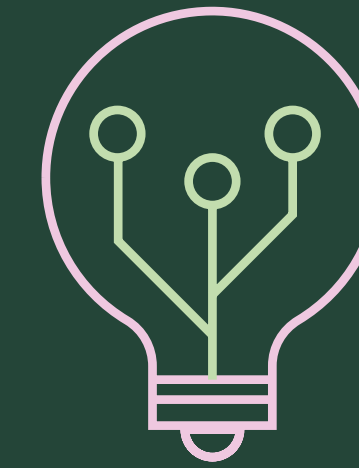
100 {
Researchers
in total



25 {
Principal
Investigators



35 {
PhD and
Postdocs



22 {
Research
seminars



45 {
Projects



65 {
Publications

Partners



UNIVERSITY
OF OSLO



UiT The Arctic
University of Norway



NR

**NORSK
REGNESENTRAL**
NORWEGIAN
COMPUTING CENTER



Norwegian
Centre of
Excellence

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 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

admin@integreat.no
integreat.no

Follow Integreat on [LinkedIn](#)

integreat