Expert matching at RCN
– from evaluation to experimentation

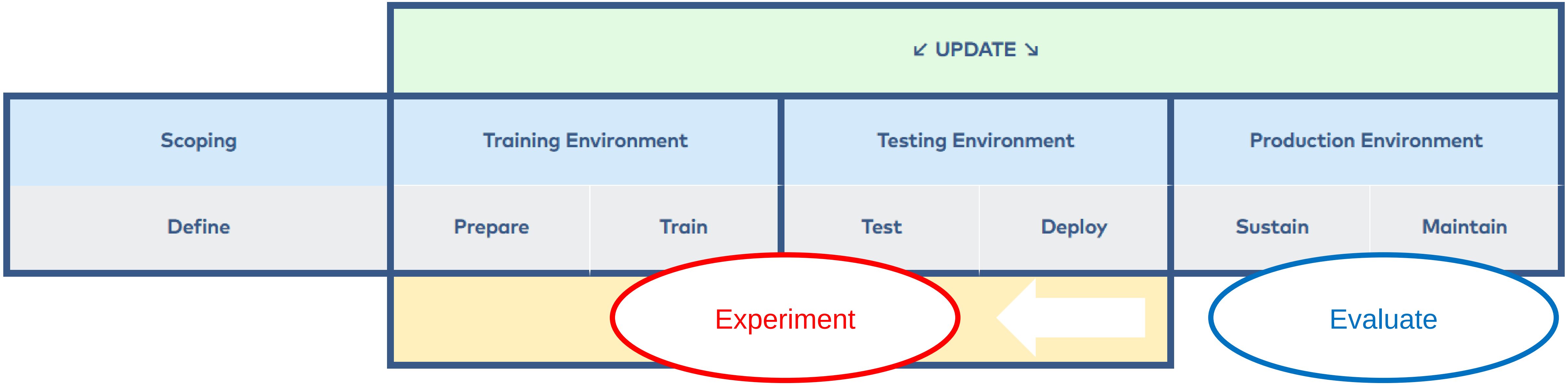
AFIRE AI sprint GRAIL workshop
24 June 2025

Special adviser Jon Holm
Data science team lead Marit Ø Gjersdal

The Research Council of Norway

AI lifecycle process flow

The AI lifecycle describes the different phases from the inception of AI systems to operation and is an iterative process



The **Definition phase** is important for scoping the intended use and setting internal requirements (for, e.g., performance and availability), but also for making sure potential fairness issues (if applicable) will be addressed properly.

The **Training environment** contains data management, data preparations, and model building activities (including algorithm selection, training, and optimization).

The role of the **Testing environment** is to test data quality, model performance, prediction uncertainty, to make decisions whether to deploy or not, and to deploy models for production if tests pass.

The **Production environment** contains activities required for flawless operation; monitoring, logging, model update logic, drift detection & countermeasures, provision of robust inference service, and mechanisms for failure recovery

Development process for RCN AI/ML team

Phase	Milestone	Responsible	Result
Start	**Submission** of needs from the business Clarify priorities with the product owner	Team Leader	**Description of needs** using user stories, etc. **Cost/benefit** assessment
Start	**Decision** on concept development	Product Owner	Project created in **Jira.** Planning documents in **Confluence**
Concept	**Development Lead** appointed by Team Leader (responsible for epic)	Team Leader	Development lead describes **function, available data,** and creates an **initial cost estimate** for the development of the service
Concept	**Decision** to start development	Product Owner	**For larger projects:** Costs accepted by the department director
Risk and Effects	Planning and conducting a **workshop for Responsible AI**	Evaluation Lead	Relevant parts of the organization are involved in assessing the risk associated with different types of use of the service and its results
Risk and Effects	**Assessment of risk, intended effects, and possible unintended effects**	Product Owner	Report on **anticipated effects and risk mitigation**
Technical Development	**Model development** and testing **Documentation** for the service	Development Lead	Detailed planning in Jira. Completed **Data Sheets** and **Model Cards** Documentation with guidance for use and **plan for follow-up**
Technical Development	**Final approval** report to oversight committee	Product Owner	Updated plan for **periodic follow-up** of services: Monitoring, evaluation, and retraining

Tools for process and documentation

Impact analysis

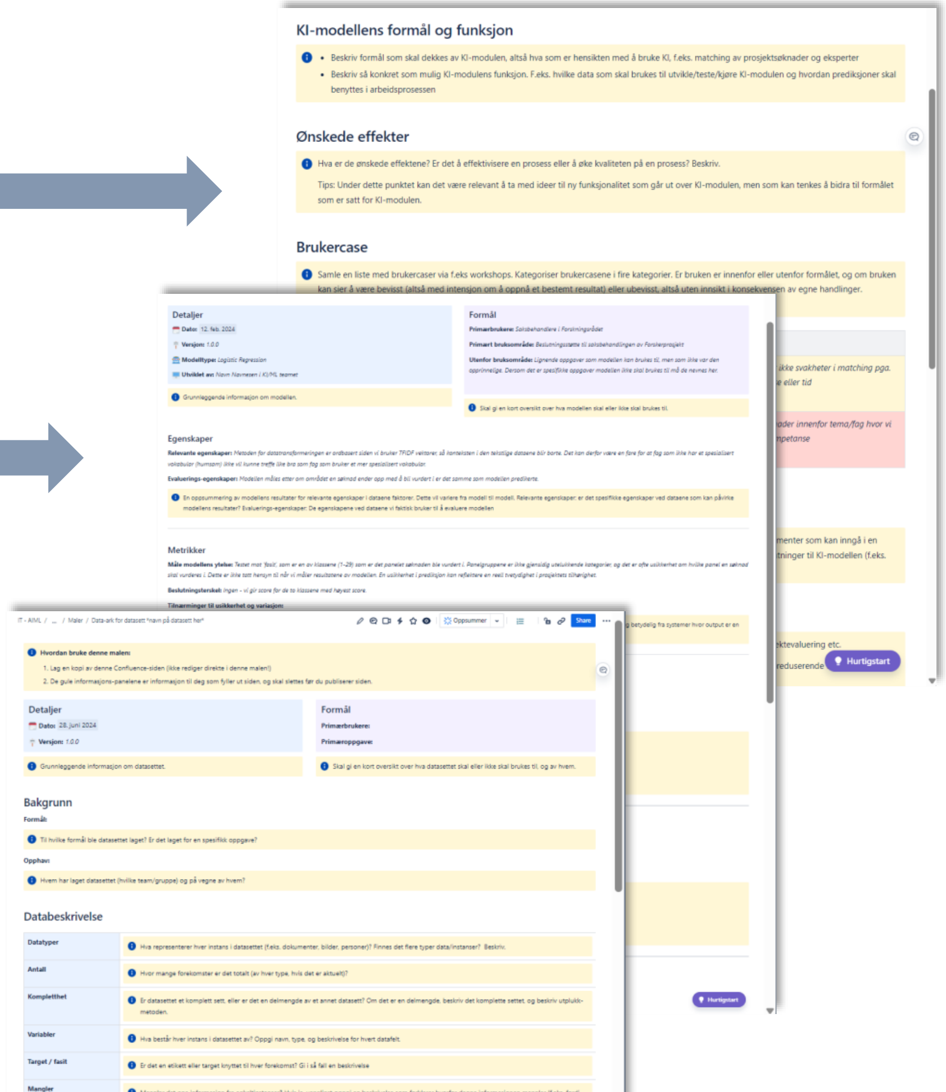
- Spengeman, Aoife, et al. "Wellcome Data Labs: An Ethics Approach to Data Science Product Development-a Intentional Work in Progress." *Available at SSRN 3974121* (2021).

Model cards

- Mitchell, Margaret, et al. "Model cards for model reporting." Proceedings of the conference on fairness, accountability, and transparency. 2019.

Datasheets

- Gebru, Timnit, et al. "Datasheets for datasets." *Communications of the ACM* 64.12 (2021): 86-92.



Follow-up of AI Services at RCN

Phase	Milestone	Responsible	Result
Monitor	**Performance data collected**	Development Lead	**Technical precision checked**
Evaluate Use & Effects	**Data on use** and/or effects on caseworkers and external stakeholders collected	Evaluation Lead	Use of the service evaluated against 'guidance for good use’ **Reassessment of the AI service as a whole**: Does it still meet the need? Is it fit for purpose?
Retrain	**Training data updated** **Model retrained**	Development Lead	**Updated documentation**

Evaluation of expert matching
DATA – MODELS – USE – EFFECTS

… and thoughts on future experiments

DATA

- How are experts identified and **recruited**?
- How is the **competence** of relevant reviewers represented in the data today?
  RCN: Publications in Scopus
- What are the known **biases** in the data sources used?
- What **other data** on reviewers' competency and appropriateness would be useful?

=>

Intervention (data): Variation in data representing expert competence

Outcome (model): Performance of model (see next slide)

MODEL

- How is **model performance** measured?
- What are the types of considerations for choosing between models?
  - **Benchmarking** towards manual selection (ground truth)
  - **Uniformity** of results across disciplines (minimise bias)
  - **Transparency** and explainability
  - **Privacy** concerns
  - **Computational cost** / environmental impact
  - …
- Which **information** does the case officer receive with the prediction?

=>

Intervention (model) : Change in AI-model

Outcome (model) : Performance measures

USE

- Do you **monitor use**? i.e. Thumbs up/down if the suggestion was useful or otherwise
- What data and knowledge is used by case-officer to **check the prediction**?
- What reasons are there for the **experts not being used**?

- What is the **response rate** of the predicted experts?
- What is the **share of precited experts** among all used experts?
- Does the prediction of experts serve **other purposes** than matching her to a proposal?
- Does the matching algorithm contribute to **good practice** in the assignment and use of experts other than speeding up the process?

=> Experiment
Intervention (model): Change information provided to case officer (availability, specialisation)
Outcome (use) : Performance measures (response rate of predicted experts)
Thumbs up/down

EFFECTS

- How can the prediction of experts **affect funding outcomes**?
  - By altering the pool of experts and their assignment to proposals: Some experts are more predicted and are used more often.
  - Can highly matched reviewers create a feedback loop?

- **Efficiency**
  - How much time do you spend recruiting experts (compared to before)?
  - Are there other changes influencing the efficiency of the review process?
  - How often do you go beyond the experts predicted by the system?
  - How many experts do you have to ask in total to recruit one expert?

=>

Intervention (model) : variation in information provided to case officer with the prediction

Outcome (effects) : time spent on recruiting experts

Scope for cooperation

1. Use some of the same questions
   in **evaluations** of existing matching algorithms

   RCN: Automn 2025

2. Test some of the same hypothesis
   in **experiments** with new matching algorithms

   RCN: Automn 2026

Empowering ideas for a better world

The Research Council of Norway