

LLMs and machine learning in future-oriented collective intelligence

**Carine van Oosteren, Anastasia Yagafarova,
Jasper van Kempen, Amber Geurts**

The Signal in the Noise

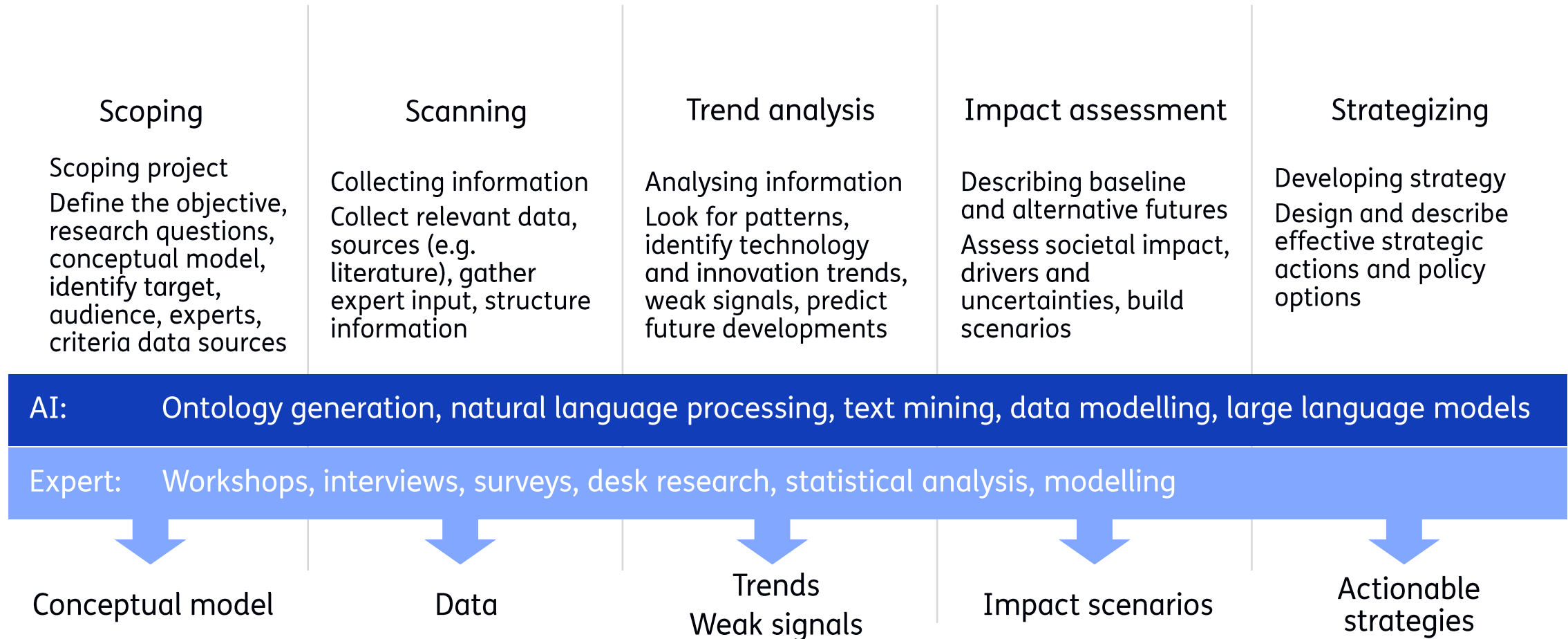
- Do you know how many policy documents on innovation policy there are?

The Signal in the Noise

- Do you know how many policy documents on innovation policy there are?

-> Lots of unstructured data

Foresight framework and the integration of AI with experts



Example 1

	Safe and resilient society	Agricultural transition and water	Climate adaptation and biodiversity	Healthcare and well-being	Circular economy and resource management	Energy transition and mobility
Reduce dependence on Asian/American semiconductors	e	e	e	e	e	e
Reduce dependence on Chinese critical minerals	e	e	e	e	•	e
Lower AI implementation costs; strengthen European AI innovation	e	e	e	e	e	e
Strengthen the European battery industry through shorter supply chains					○	•
Maintain EU leadership in cleantech despite Chinese competition						•
Develop hydrogen production as an alternative to fossil fuels						•
Strengthen domestic defence industrial capacity	•					
Maintain autonomous access to space	•		•			
Maintain pharmaceutical leadership; strengthen position in dynamic segments				•		
Enhance cybersecurity against high-risk suppliers	•			e		e
Reduce dependence on foreign telecom/5G/6G	e			e		e
Strengthen cloud sovereignty to ensure EU data control	•			e		e
Increase independence in quantum technologies	e	e	e	e	e	e
Strengthen biotechnology innovation		•	○	•	•	
Build capacity for advanced materials	e	e	e	e	•	e
Strengthen the circular economy for raw material use					•	
Transition to clean energy to reduce fuel dependency						•
Develop charging infrastructure for alternative fuels						•
Improve cross-border energy infrastructure						•

Example 2

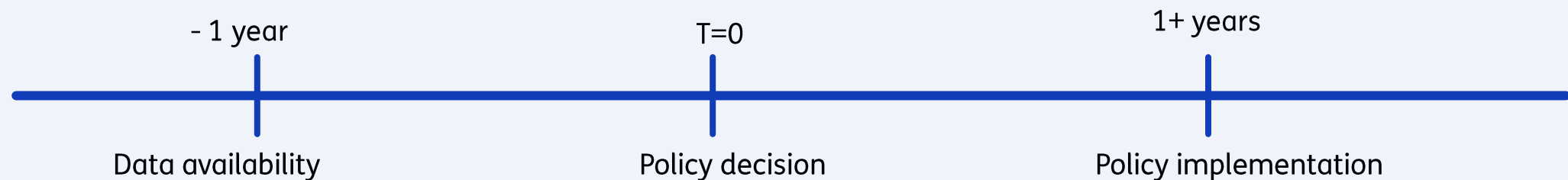
NOWCASTING

FOR BETTER DECISIONS
NOT ONLY IN
TIME OF
CRISIS

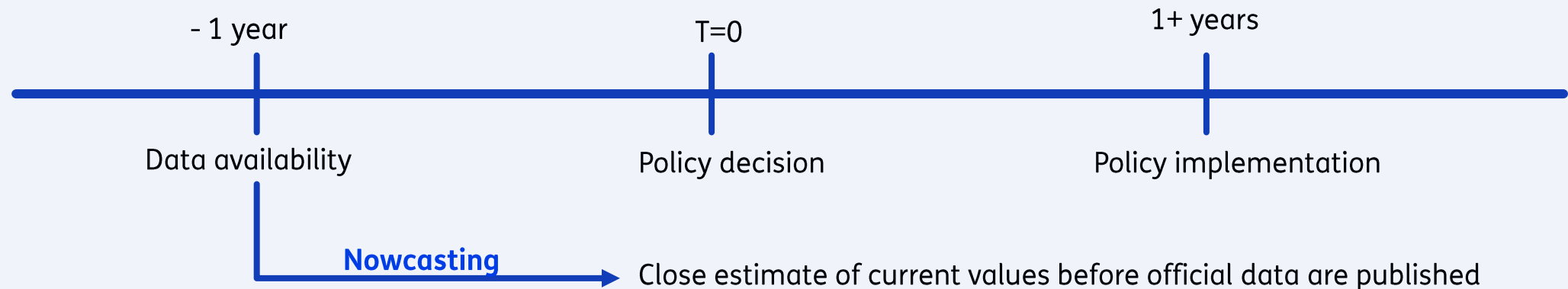


Polymakers often rely on outdated statistics

- Many indicators, like R&D expenditures, have at least a 1 year delay; other indicators, related to growth of the digital economy: 1-4 years delay.
- Limits governments' ability to make timely, informed decisions



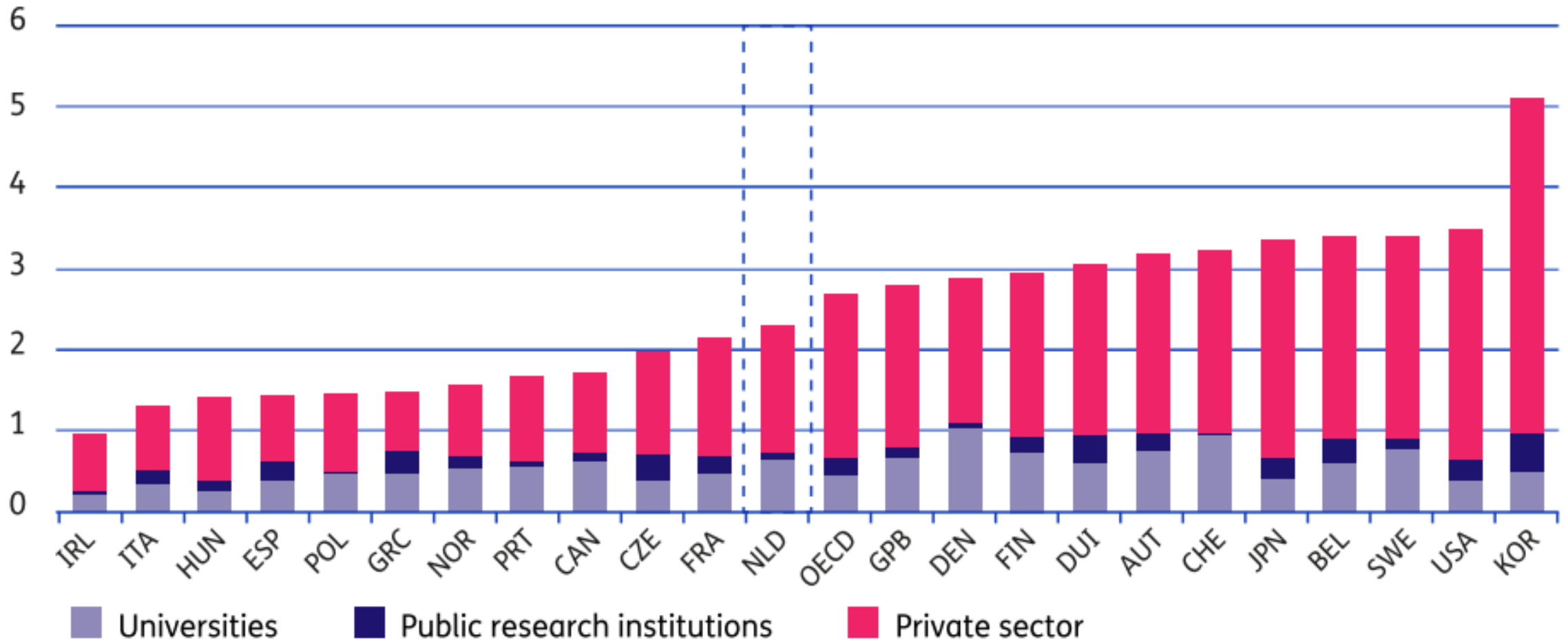
How can AI nowcasting help Policy Making?



Conventional approach
Time series & linear regression

Recent work
Machine learning model: use of **high frequency Google Trends Data** and a **Neural Network** for more sophisticated nowcasting

R&D expenditure as % of GDP, 2022



R&D expenditure of top 10 companies The Netherlands, 2023



- Nowcasting opens the door to faster, cheaper and more frequent tracking of R&D investment data
- It can be used in:
 - Yearly publication R&D Top 30
 - Action plan 3% R&D in 2030 (work in progress)

How do we nowcast R&D?

We use the EPFL-method

The model predicts annual R&D spending using a mix of:

- Past R&D values
- Economic indicators (e.g. GDP, employment)
- Google search trends on topics like “start up funding” or “innovation strategy”

Data used for 8 countries: Switzerland, US, UK, Germany, South Korea, Japan, China and Canada

Ref: Aboutorabi, A., & de Rassenfosse, G. (2024). Nowcasting R&D expenditures: A machine learning approach. [\[2407.11765\] Nowcasting R&D Expenditures: A Machine Learning Approach](#)

Stakeholders and their respective search terms for R&D expenditure

Stakeholder	Search Terms
Businesses	R&D Expenditure, Product Development
Consulting Firms	Innovation Management, Innovation Strategy
Government Agencies	Government Grants, Research Funding
Innovation Hubs	Startup Incubation, Technology Park
Patent Attorneys	Patent Attorney, Patent Registration
R&D Employees	R&D Jobs
Research Institutions	Collaboration with Industry, Research Grant
Tax Authorities	R&D Tax Credit
Venture Capitalists (VCs)	VC Investment, Startup Funding

What did we do?

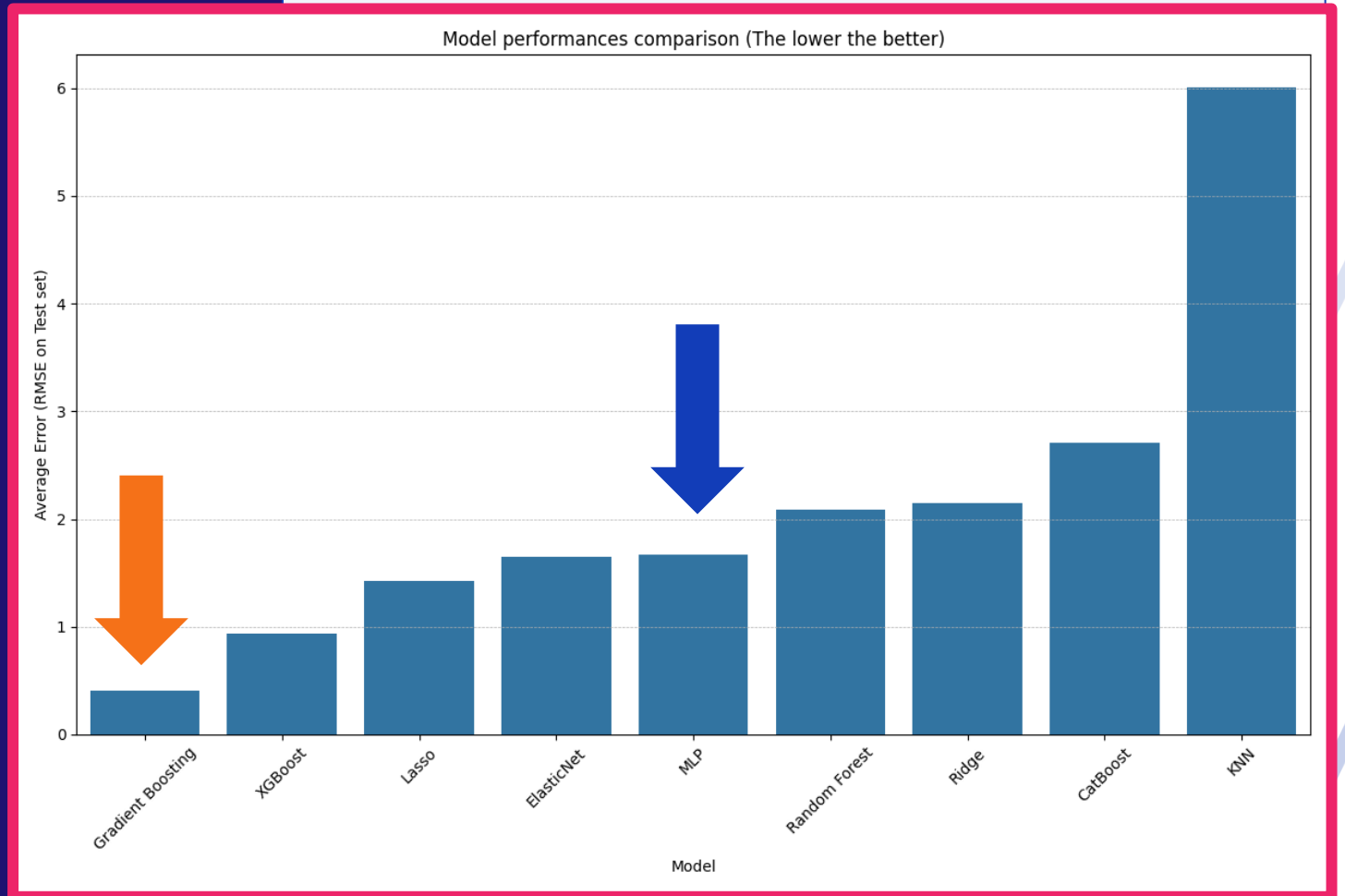
We used
their
method

We tested
additional
models

We included
data of The
Netherlands

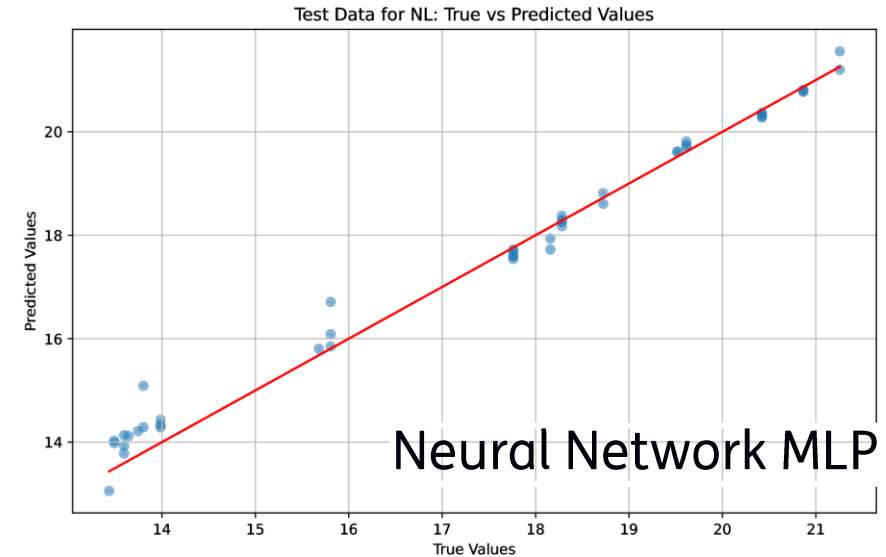
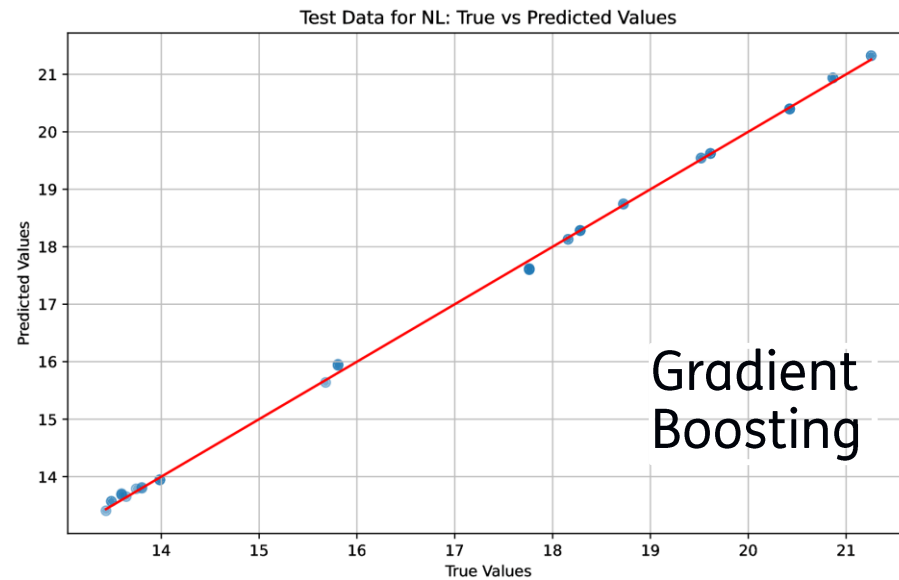
Model comparisons

- When comparing the performances of various models the gradient boosting is the best.
- MLP is the neural network from the EPFL-paper



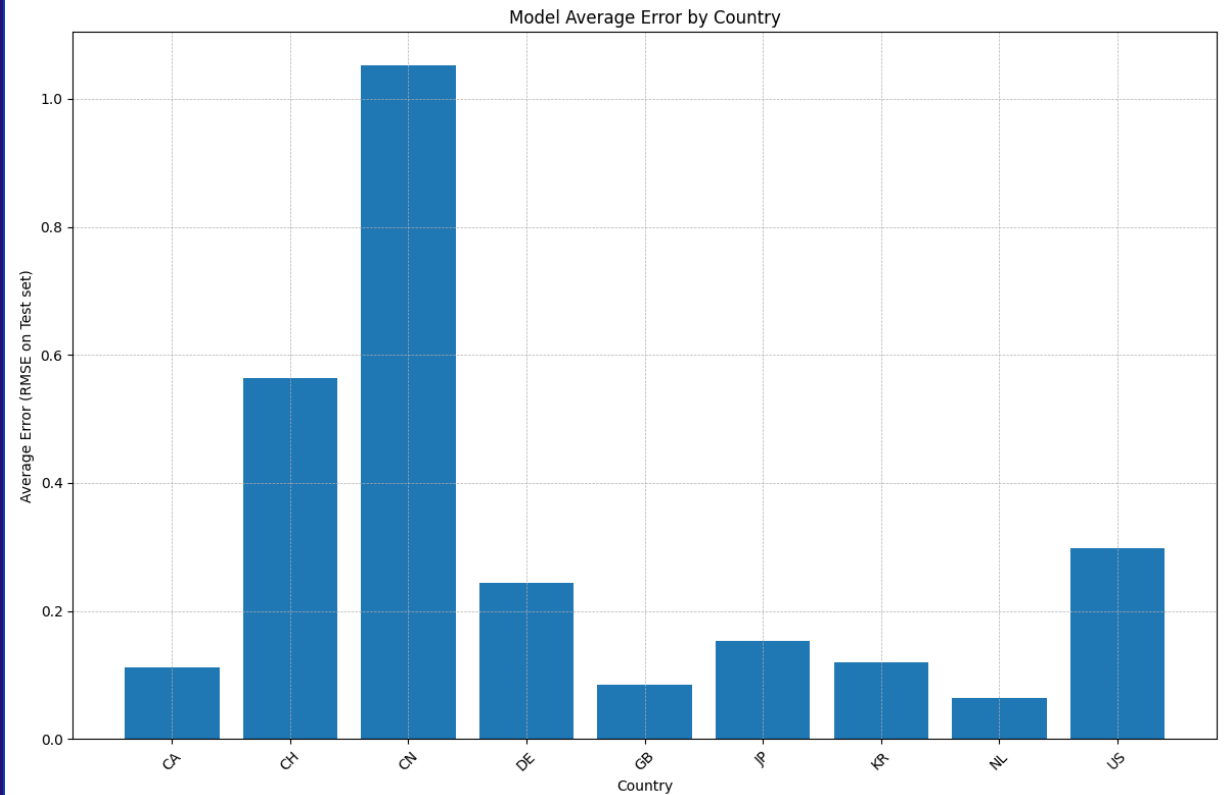
MLP versus Gradient Boost performance

For the ideal model all the dots should be on the line



Performance per country of Gradient Boosting

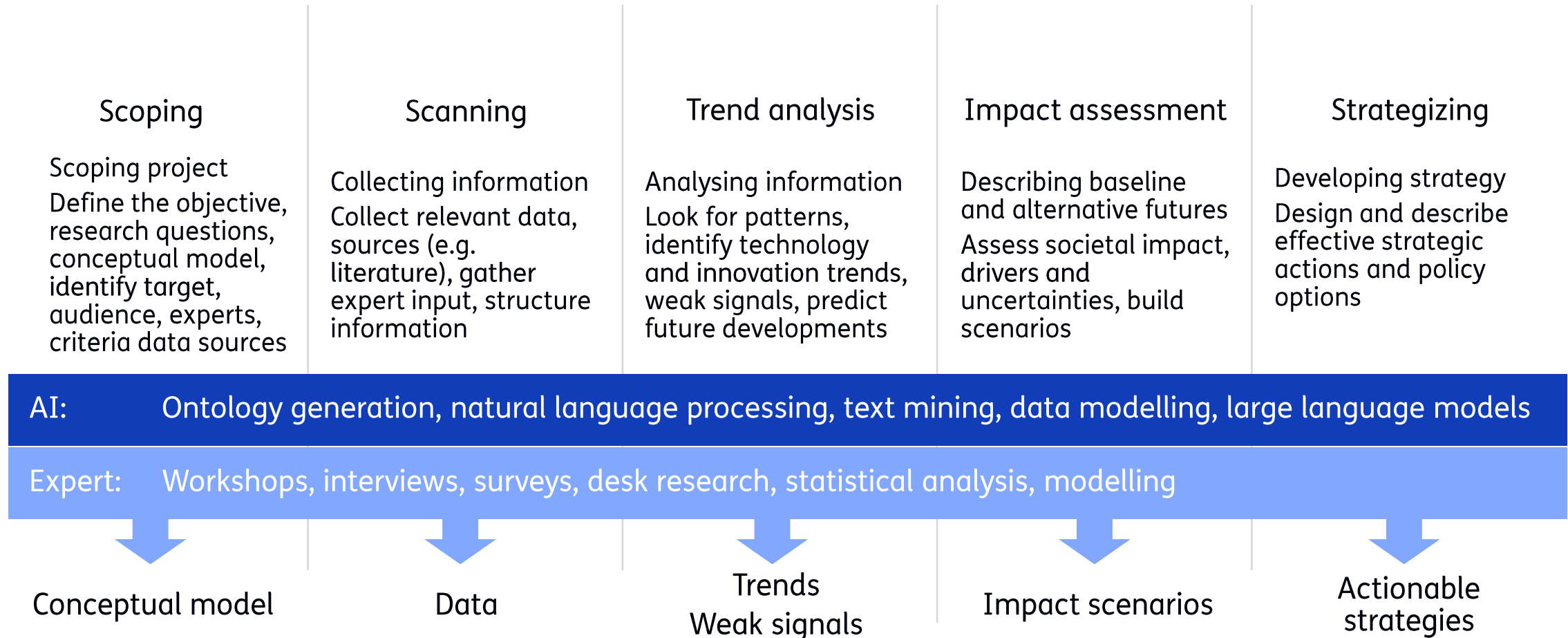
- Best performance for The Netherlands
- Worst performance for China.
- Other models have different outcomes per country



So, finally our preliminary result of Dutch R&D expenditure in 2024



Foresight framework and the integration of AI with experts



Any questions, please contact us

LLMs

Anastasia Yagafarova anastasia.yagafarova@tno.nl

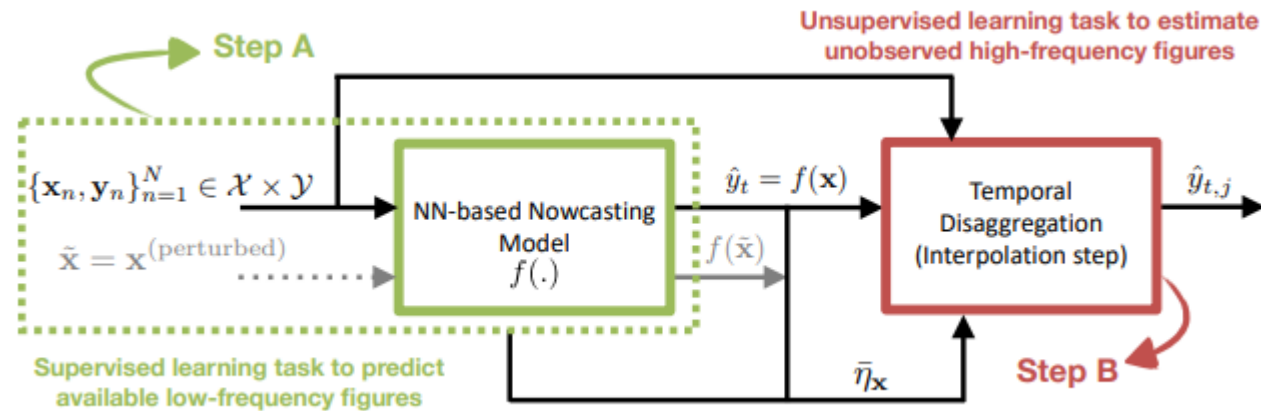
Amber Geurts amber.geurts@tno.nl

Nowcasting

Carine van Oosteren carine.vanoosteren@tno.nl

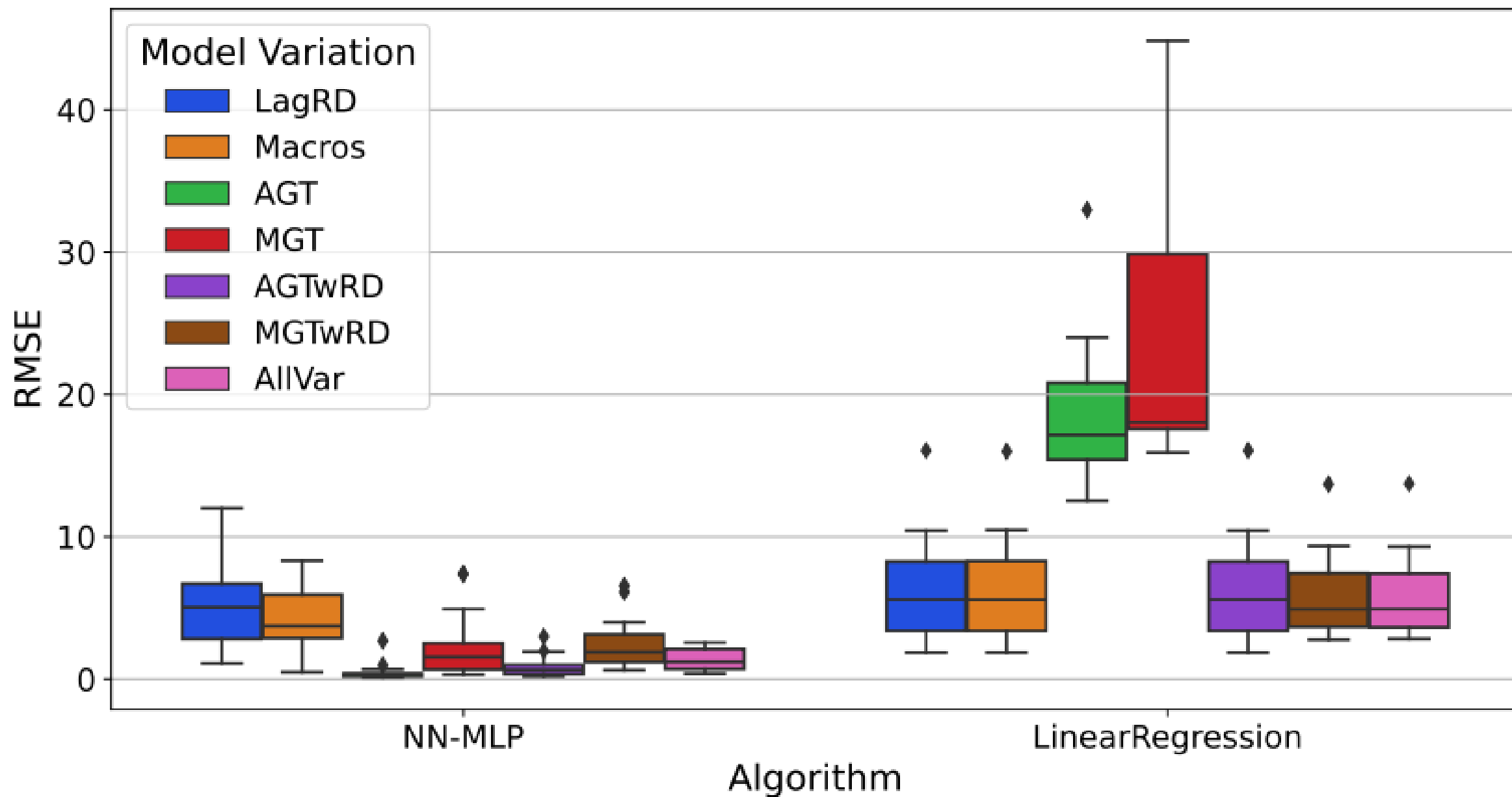
Gabriel Hoogerwerf gabriel.hoogerwerf@tno.nl

Jasper van Kempen jasper.vankempen@tno.nl (back in June)



Step B: Monthly Estimates of R&D expenditure (GERD)

Breakdown of yearly predictions into monthly estimates by analyzing how relevant search terms vary across months



Results EPFL

- The neural network using Google Trends alone is surprisingly effective—sometimes even more accurate than models using traditional economic data.
- Predictions made by the model correlated well with real-world data like employment in R&D sectors, suggesting it captures genuine economic patterns.
- The model outperformed traditional forecasting methods, especially in handling complex, high-volume data.
- Hybrid method works best: ALL Var: macro economic indicators, historical R&D and Google Trends