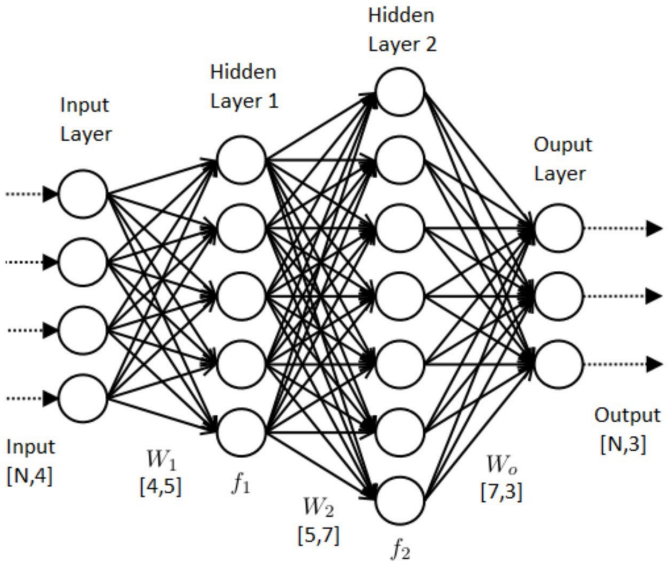
AI in clinical trials – a regulatory perspective

Gabriel Westman

Associate Professor (MD, PhD, MSc Eng)
Head of Artificial Intelligence, Swedish Medical Products Agency
Member of EMA CHMP Methodology Working Party
Member of HMA/EMA Network data steering group

AI basics

- High-parameter models capable of processing complex input data
- Machine learning – an iterative data-driven process with feedback
- White box, black box and explainable AI
- Large data sets or transfer learning needed for development



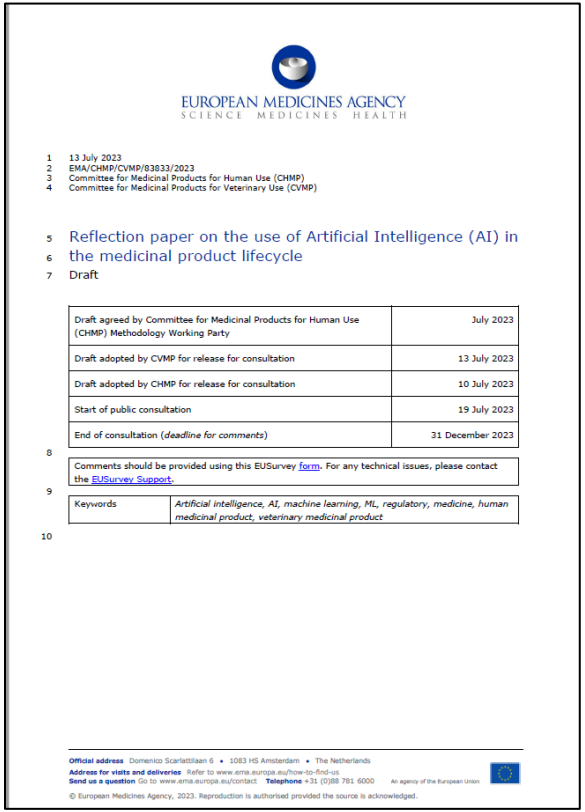
Separating the hope from the hype

- We are not getting AGI anytime soon
- Encoder models are still the dominating workhorse when it comes to delivering business value
- LLMs are easy to overestimate – beware of ”imitation of work”
- Select use cases where the output from the model is the actual value, rather than a proxy for an underlying value (such as an assessment)

EMA reflection paper on AI

- Provides considerations on the use of artificial intelligence (AI) and machine learning (ML) in the lifecycle of medicinal products
  - Describes the current experience in the EMRN
  - Acknowledges fast evolution of in the field of AI/ML
  - Should be read in coherence with both legal requirements and overarching EU principles on AI, data protection, and medicines regulation
  - Not to be considered a regulatory guidance document

- Lead: CHMP Methodology Working Party



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

13 July 2023
EMA/CHMP/CVMP/83833/2023
Committee for Medicinal Products for Human Use (CHMP)
Committee for Medicinal Products for Veterinary Use (CVMP)

Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle
Draft

Draft agreed by Committee for Medicinal Products for Human Use (CHMP) Methodology Working Party	July 2023
Draft adopted by CVMP for release for consultation	13 July 2023
Draft adopted by CHMP for release for consultation	10 July 2023
Start of public consultation	19 July 2023
End of consultation (*deadline for comments*)	31 December 2023

Comments should be provided using this EUSurvey form. For any technical issues, please contact the EUSurvey Support.

Keywords	*Artificial intelligence, AI, machine learning, ML, regulatory, medicine, human medicinal product, veterinary medicinal product*

Table of contents

AI in the medicinal product lifecycle

- AI and ML tools can - if used correctly - effectively support the acquisition, transformation, analysis, and interpretation of data within the medicinal product lifecycle.
- AI introduces new risks that need to be mitigated to ensure the safety of patients and integrity of clinical study results.
- Important differences between the human and veterinary domain include legal bases, regulatory requirements and ethical issues.

Key regulatory principles

- It is the responsibility of the applicant or MAH to ensure that all **algorithms, models, datasets,** and **data processing pipelines** used are **fit for purpose** and are **in line with ethical, technical, scientific, and regulatory standards**
- The applicant or MAH is expected to provide a scientific base along with **sufficient technical details to allow comprehensive assessment** of any AI/ML systems used in the medicinal product lifecycle, the integrity of data, and generalisability of models to the target population and specific context of use.
- While acknowledging that AI technology holds the potential to improve many aspects of the medicinal product lifecycle, **trustworthiness for regulators, payers and patients alike must not be compromised** by the introduction of new technology.

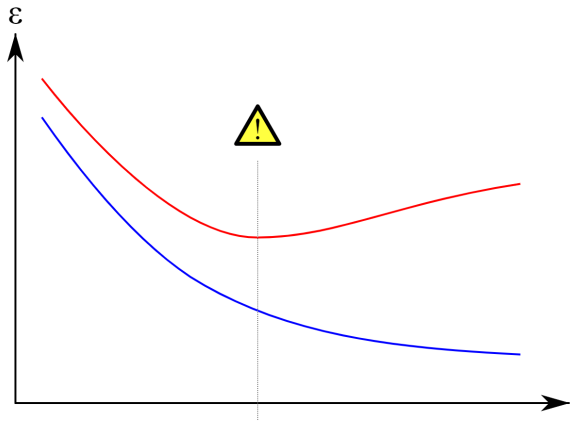
A risk-based approach

- A risk-based approach for development, deployment and performance monitoring of AI and ML tools allows developers to pro-actively define and mitigate risks throughout the AI/ML system lifecycle.
- The degree of risk depend on several factors and may vary throughout the lifecycle of the AI-system.

  Such factors include **architecture of the AI technology**, the **context of use** and the **degree of influence** the AI technology exerts.

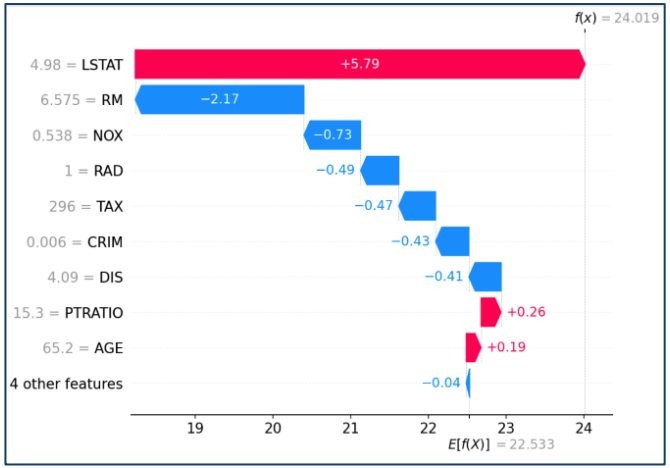
Managing overfitting and data leakage

- Depending on the level of risk/impact and context of use, the risks of overfitting and data leakage should be addressed proportionally
- For high regulatory impact settings such as in relation to the primary endpoint in late-stage clinical trials, prospective testing is expected
- For low-risk settings, testing on hold-out retrospective data may be acceptable
- Cross-validation can support internal generalisability
- Sensitivity analyses based on a calendar-time train-test data splits are encouraged



Interpretability and explainability

- Everything else being equal, the use of transparent (interpretable) models is preferred
- Use of black box models may be acceptable if needed to achieve satisfactory performance and/or robustness, but require a more rigid validation/test protocol
- The use of explainability techniques (xAI) should be used whenever possible, to provide both global and local explanations of model behavior



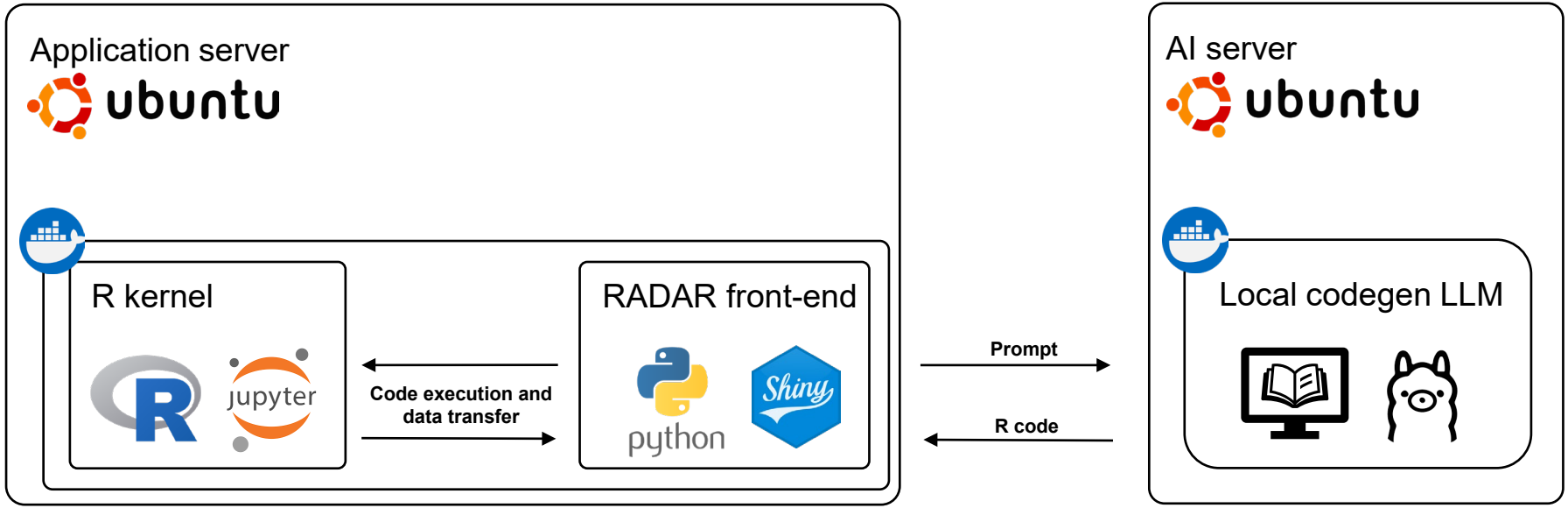
Upcoming EMA guidelines on AI

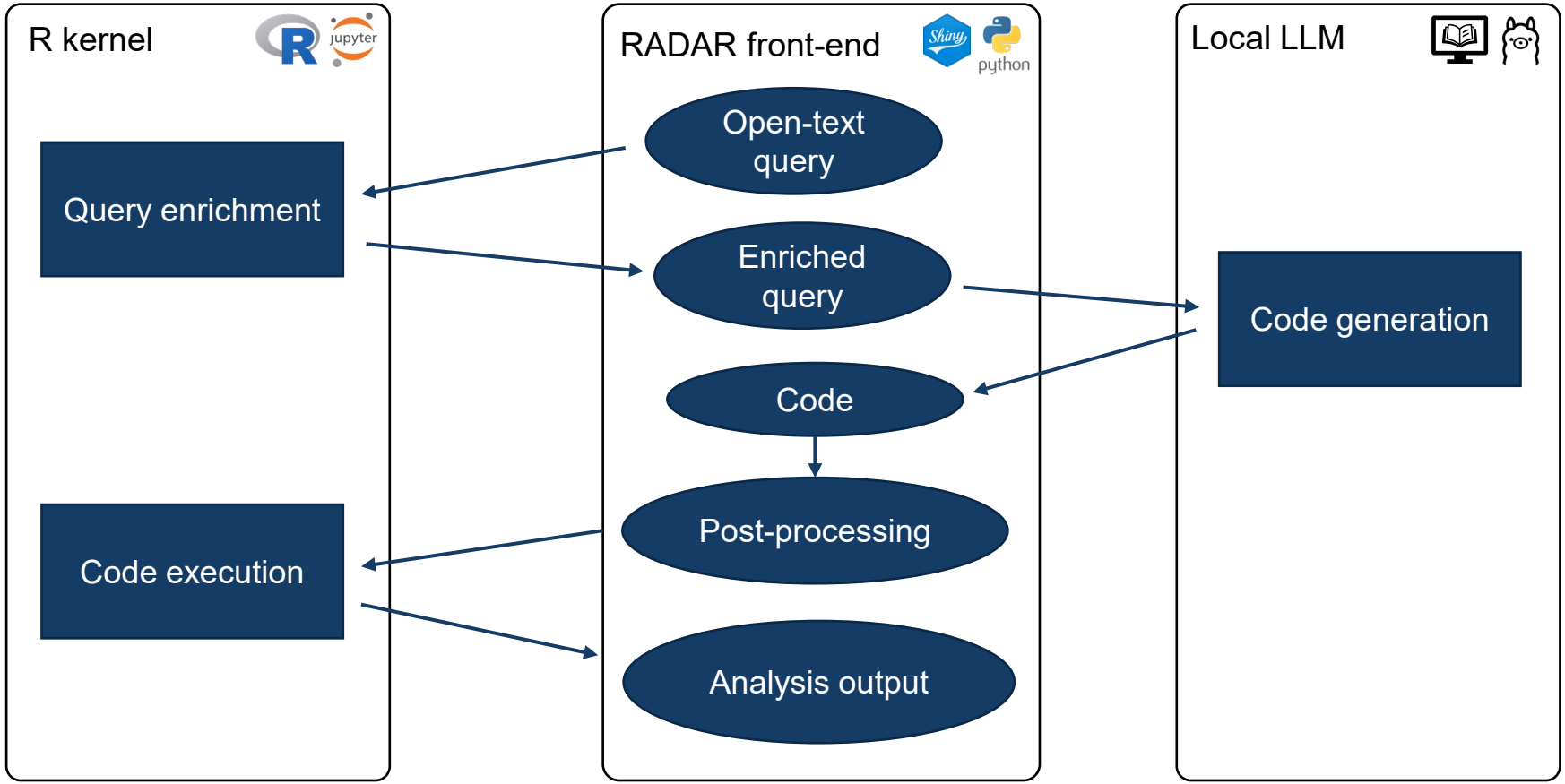
- Several comments covered topics or requests for prescriptive regulatory detail at a level that cannot be included in the format of an EMA reflection paper
- CHMP Methodology Working Party (MWP) workplan from 2025 onwards includes development of specific guidance on:
  - AI in clinical development
  - AI in pharmacovigilance

RADAR – raw data augmented review

- MPA has been actively using raw data analysis for efficacy data, clinical chemistry/haematology and pharmacometrics
- Raw data analysis is valuable in various procedural roles but needs close integration between biostatistics and clinical expertise to create value
- An AI-driven tool (RADAR – raw data augmented review) has been developed by the MPA to enable quick, no-code analysis of CDISC ADaM datasets

RADAR – system architecture




R kernel
jupyter
Query enrichment
Code execution
RADAR front-end
Shiny
python
Open-text query
Enriched query
Code
Post-processing
Analysis output
Local LLM
Code generation

RADAR - workflow

Load Data | Merge Tables | Query Data | View Data

Select Table

adqsadas_1

Explain Generated Code

Query:

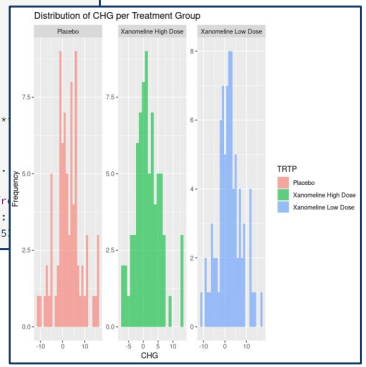
Look at the subset of data where EFFFL="Y" and ANL01FL = "Y" for AVISITN = 24 and perform a linear regression for how CHG depends on TRTPN. Print the summary of the model so that P-values are shown. Treat TRTPN as a continuous value. Also plot the distribution of CHG for this subset per TRTP specifically (different colours per category).

Generate and Execute | Generate

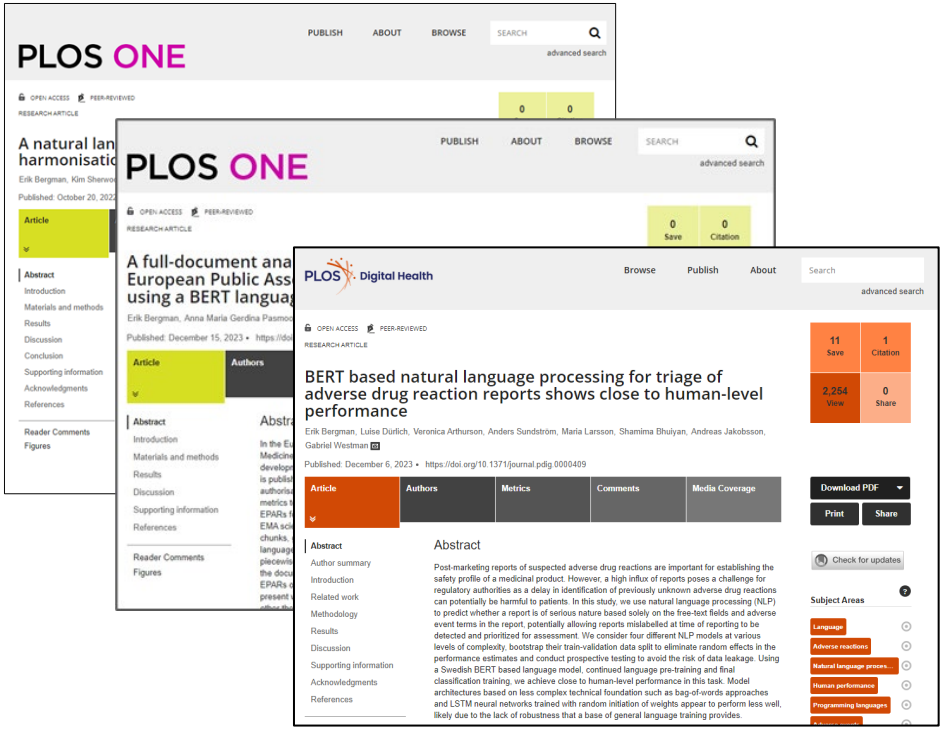
Code:

```
# Perform linear regression
model <- lm(CHG ~ TRTPN, data = subset_data)

# Print the summary of the model
summary(model)

# Plot the distribution of CHG for this subset per TRTP specifically (different colours per
category)
library(ggplot2)
ggplot(subset_data, aes(x = CHG, fill = TRTP)) +
  geom_histogram(binwidth = 1, alpha = 0.6, position = "identity") +
  facet_wrap(~TRTP, scales = "free") +
```

```
Call:
lm(formula = CHG ~ TRTPN, data = subset_data)

Residuals:
     Min       1Q   Median       3Q      Max
-13.5803  -3.3827  -0.5465   3.1089  15.1089

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.58035    0.57047   4.523 9.72e-06 **
TRTPN       -0.01276    0.01027  -1.243    0.215
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.

Residual standard error: 5.261 on 232 degrees of fr
Multiple R-squared:  0.006611,  Adjusted R-squared:
F-statistic: 1.544 on 1 and 232 DF,  p-value: 0.215
```



AI for medicines regulation: Scientific output 2022-2024



Thank you for your attention.

Credits: Erik Bergman, Luise Dürlich, Seamus Doyle, Victor Lindeman, Samuel Fransson

Contact: gabriel.westman@lakemedelsverket.se