AI in medicinal products

Liam Childs
Artificial Intelligence and Big Data Working Group

Bundesinstitut für Impfstoffe und biomedizinische Arzneimittel
Federal Institute for Vaccines and Biomedicines

Das Paul-Ehrlich-Institut ist ein Bundesinstitut im Geschäftsbereich des Bundesministeriums für Gesundheit.

The Paul-Ehrlich-Institut is an Agency of the German Federal Ministry of Health.

Paul-Ehrlich-Institut

Disclaimer

The views expressed in this presentation are the personal views of the presenter(s). They shall not be understood or cited as opinions of the Paul-Ehrlich-Institut. The presenter has not received any funding or grants from companies or from associations representing companies.

Paul Ehrlich Institute

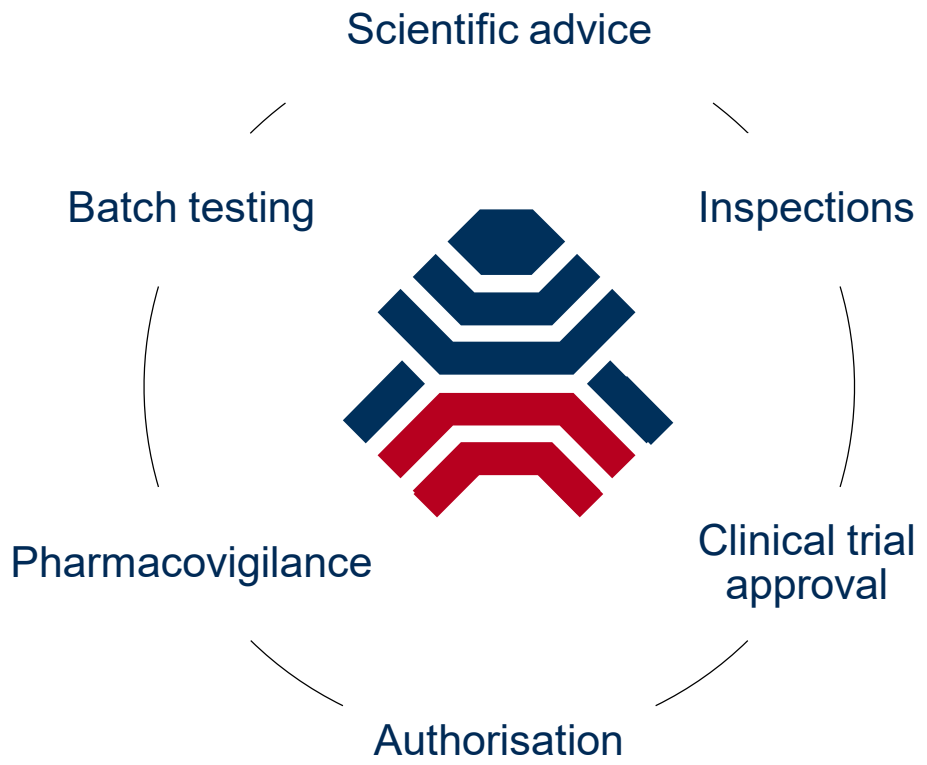
Paul-Ehrlich-Institut

Medicinal products within our remit

Vaccines	Antibodies	Advanced Therapy Medicinal Products
Blood products		Allergens
Tissue preparations	Stem cell preparations	Immunological veterinary medicinal products

Not medical devices (please see BfArM)

Duties



Outside

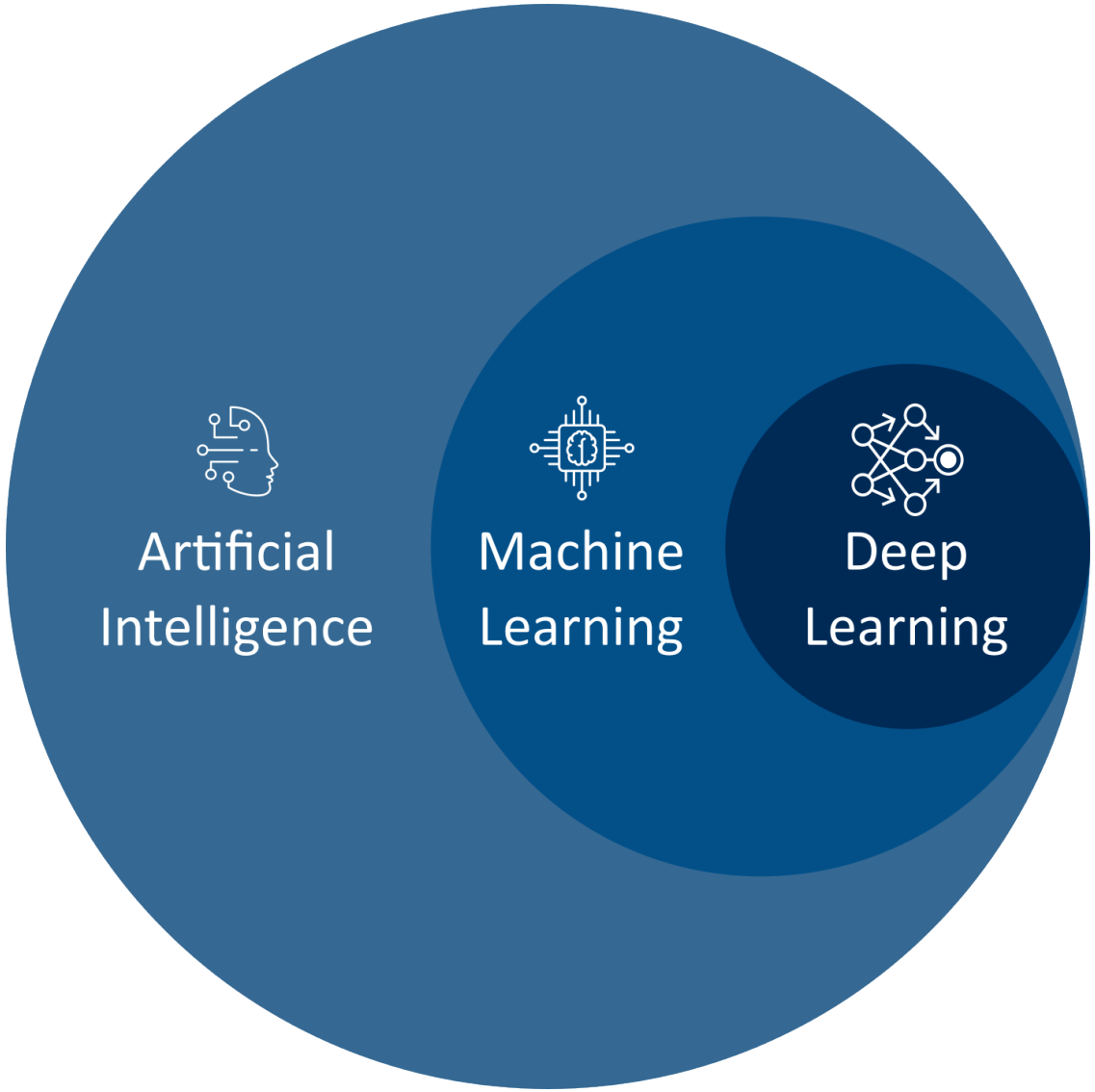
Inside

INTRODUCTION TO AI

What is artificial intelligence?

"AI is the simulation of human-like intelligence using computer systems"

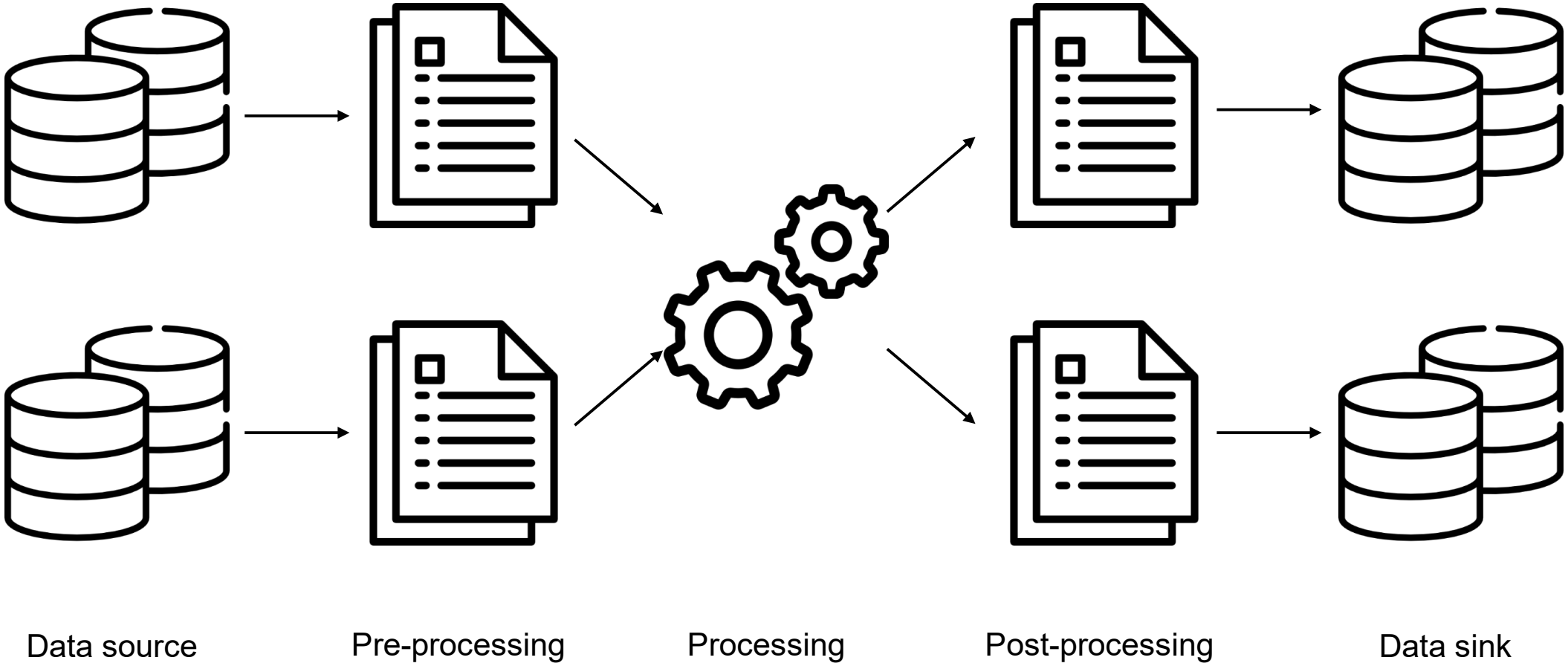
"AI is the use of computers to perform tasks that require human-like intelligence"



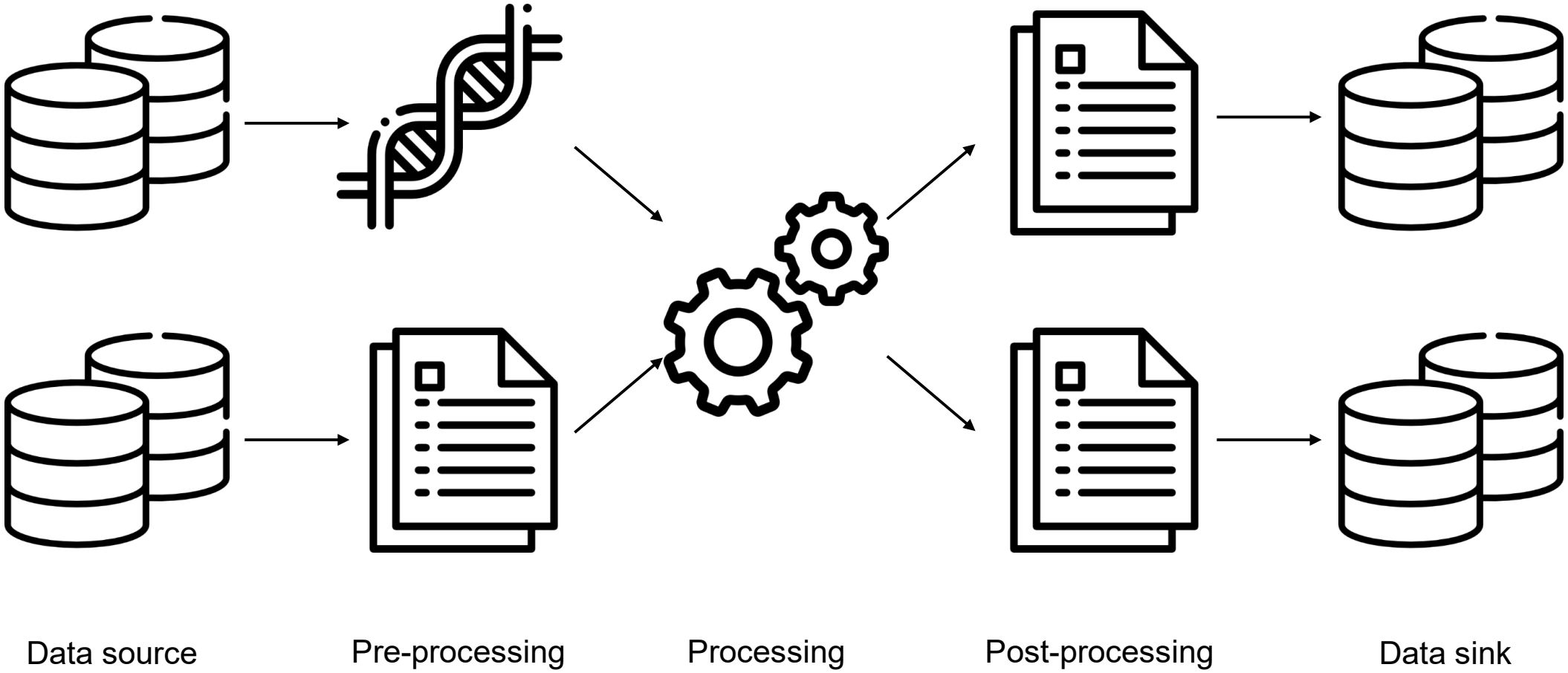
"AI combines computer science and robust datasets to enable problem solving"

"Artificial intelligence (AI) is the intelligence of machines or software, as opposed to the intelligence of humans or animals."

A step: the atomic unit of data science

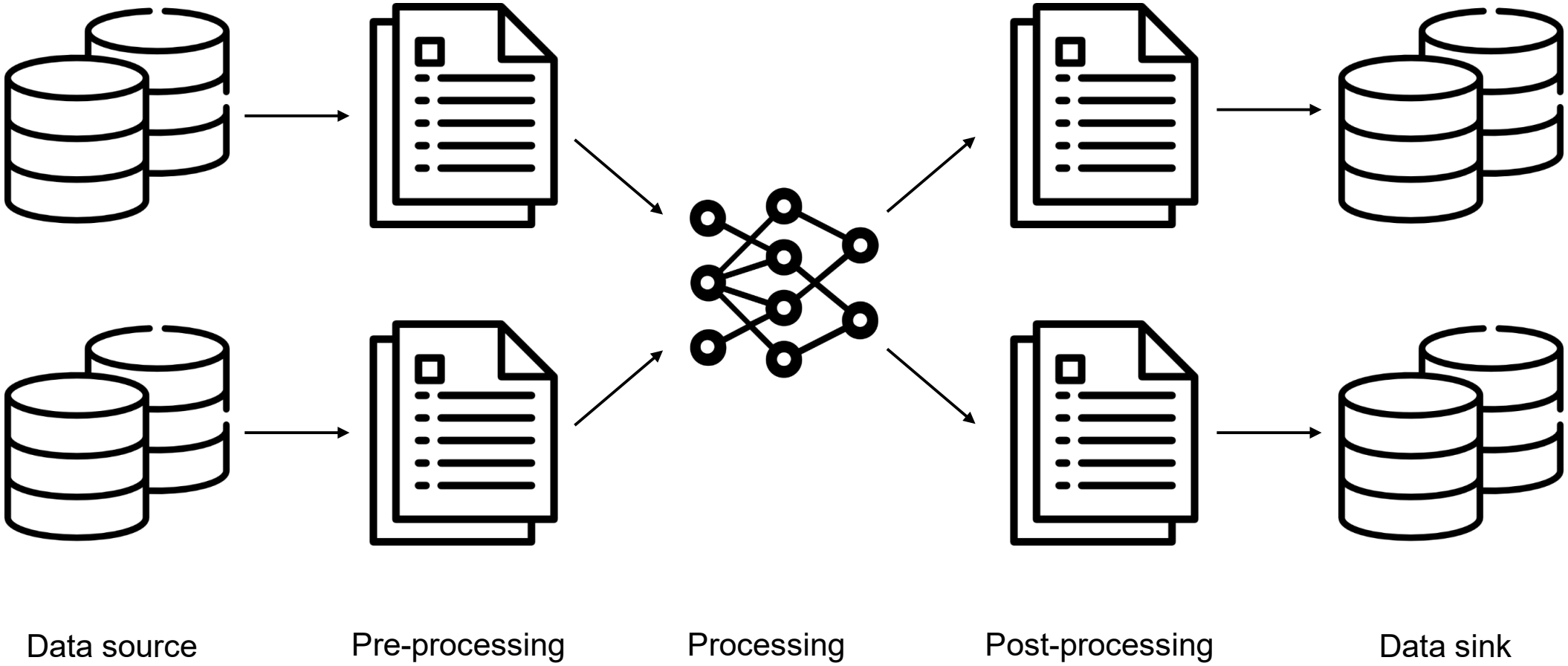


Bioinformatics: Data science using biological data



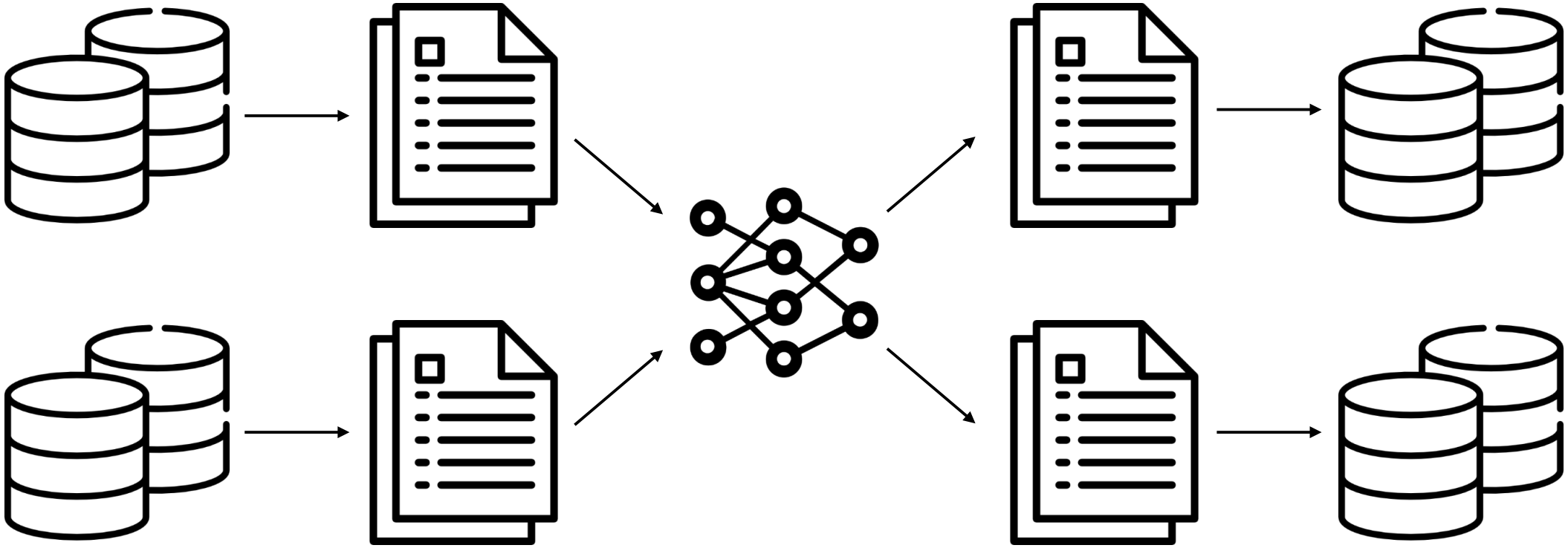
AI: Data science using machine learning

Paul-Ehrlich-Institut



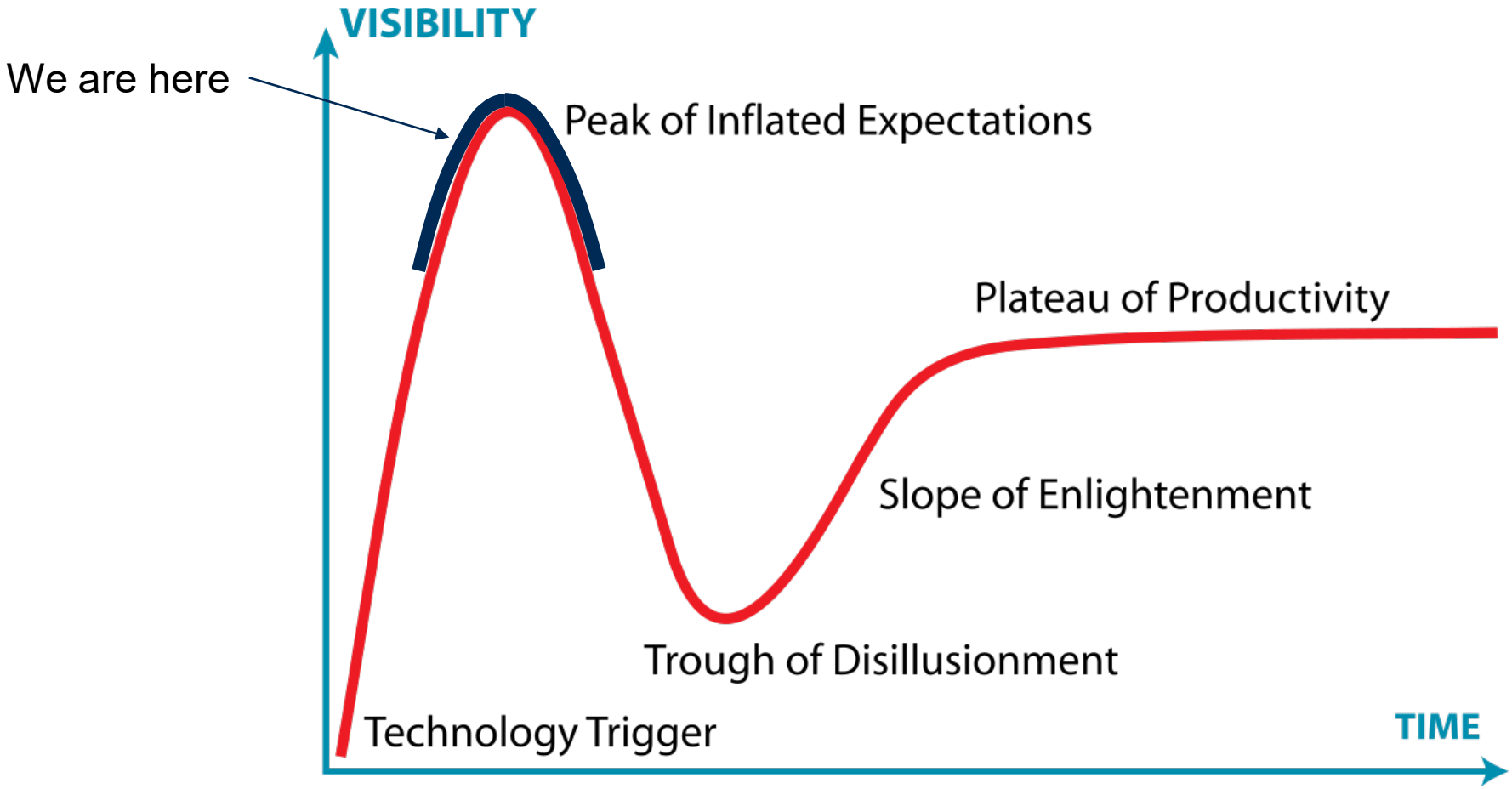
AI: Data science using machine learning

Paul-Ehrlich-Institut



AI is more than just the *processing* step

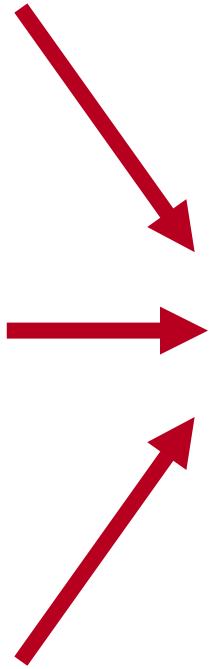
The hype around AI

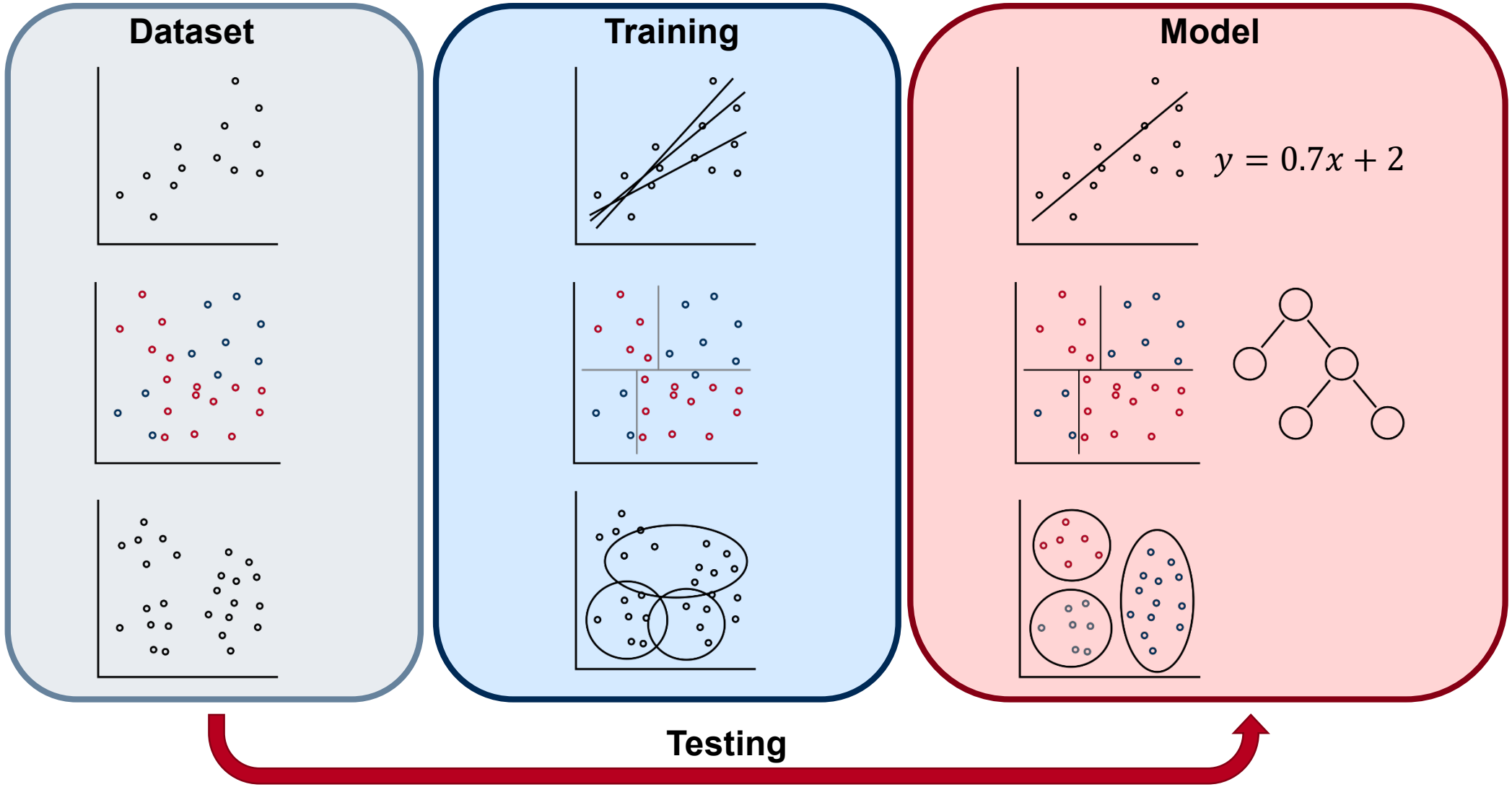


What is machine learning?

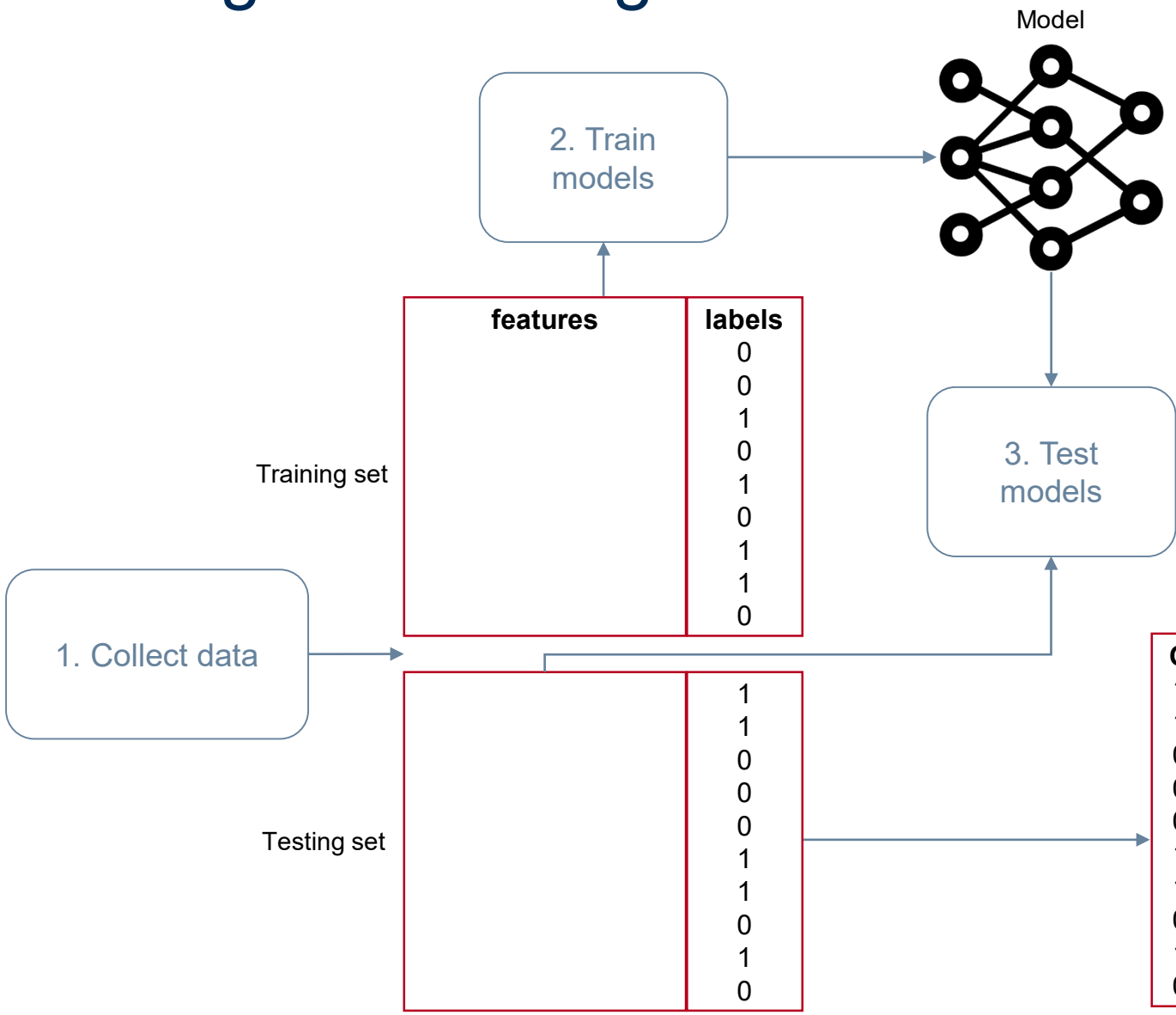
Data gathering

- Measurements
- Outcomes





Training and Testing



Features	Measurements that lead to an outcome (e.g. peptide sequence, HLA type)
Labels	The observed outcome (e.g. binding affinity)

Observed \ Predicted	**0**	**1**
0	TN	FP
1	FN	TP

	0	**1**
0	3	2
1	1	4

Confusion matrix

Performance measures

Observed \ Predicted	0	1
0	TN	FP
1	FN	TP

Precision (P)	$TP/(TP + FP)$	How many of the positive predictions are true?
Recall (R)	$TP/(TP + FN)$	How many of the observed positives did we predict?
F1-Value	$2PR/P + R$	How well are precision and recall balanced?

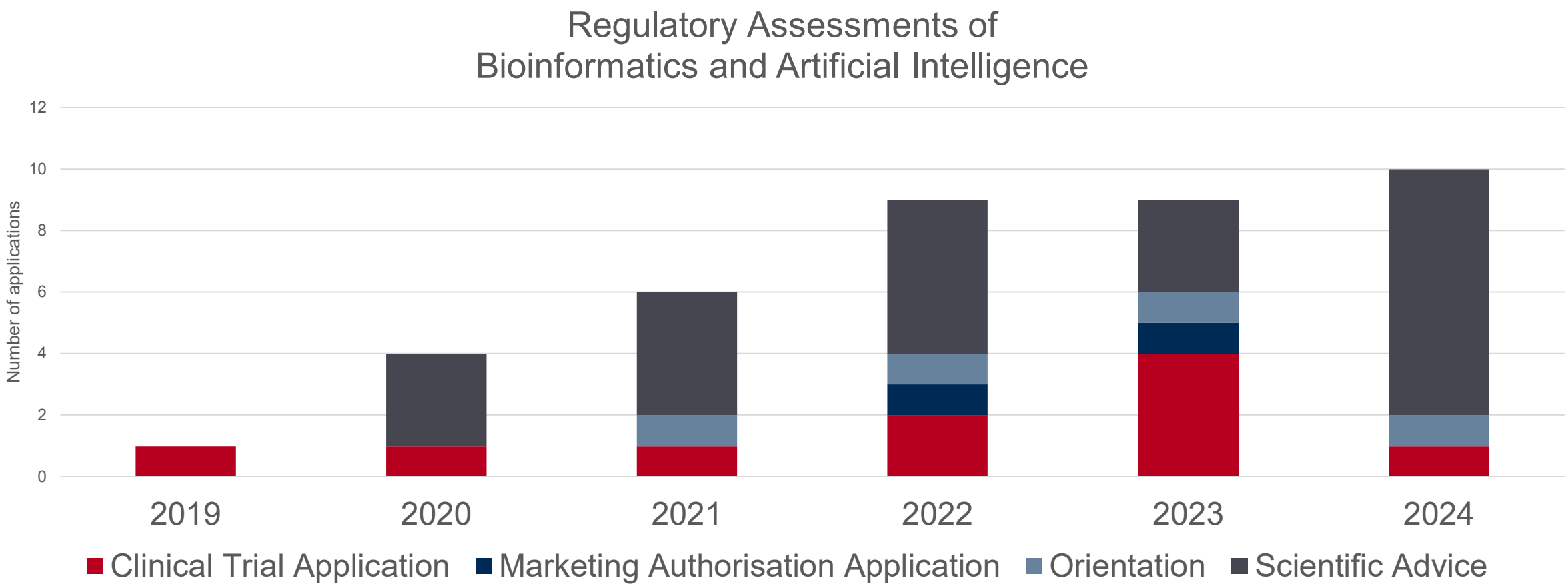
All measures range between 0 and 1, where 1 is the best performance.

Performance measures come in pairs (e.g. sensitivity/specificity).

Optimising one measure usually requires compromises on the other.

Algorithms are never perfect (you should rarely see a „1“).

Assessments at the Paul Ehrlich Institute



AI IN MEDICINAL PRODUCT LIFECYCLE

Medicinal Product Lifecycle

Research & Development

- Target Discovery & Validation
- Lead Identification & Optimisation

Pre-Clinical Studies

- Precision medicine
- Pharmacokinetics
- Pharmacodynamics

Clinical Trials

- Patient Recruitment, Selection and Stratification
- Dose Optimisation
- Adherence and Retention
- Site Selection
- Data Analysis

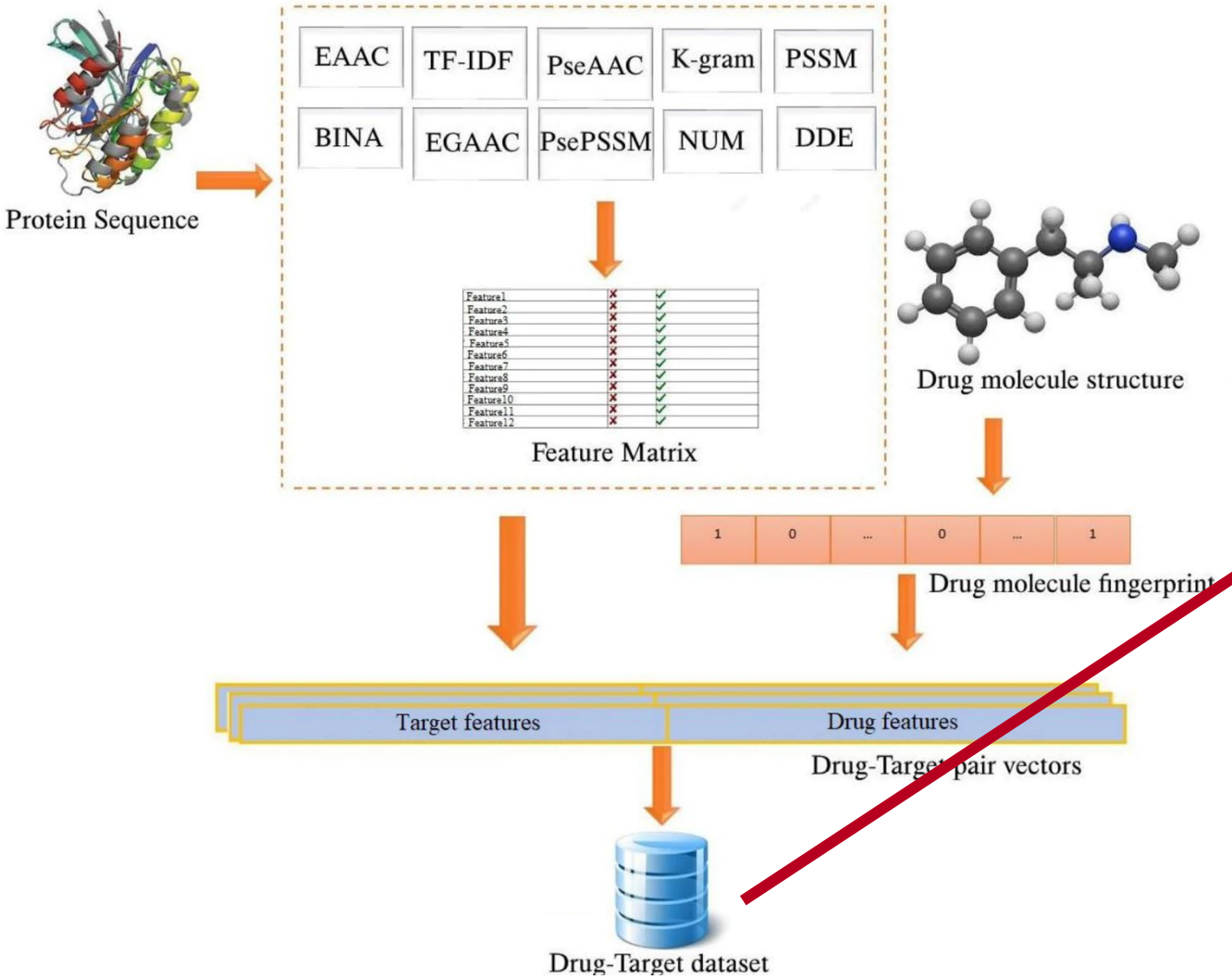
Manufacturing

- Process Control and Optimisation
- Quality Control
- Trend Monitoring

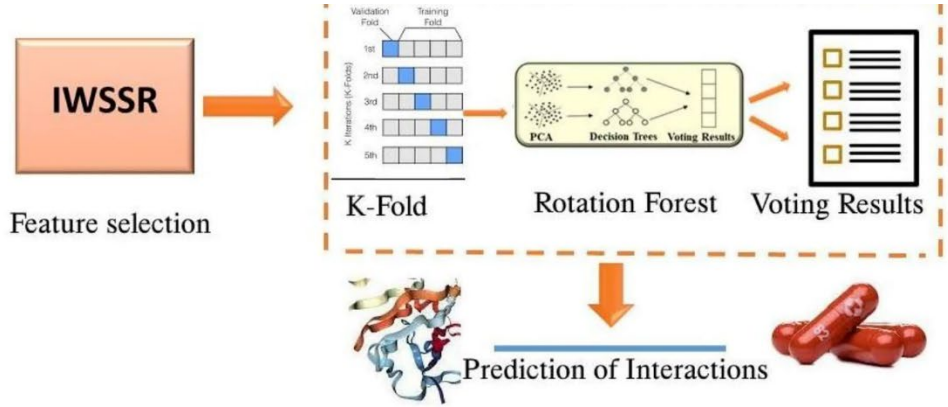
Post-Marketing Surveillance

- Risk Assessment
- Signal Detection

AI in drug discovery

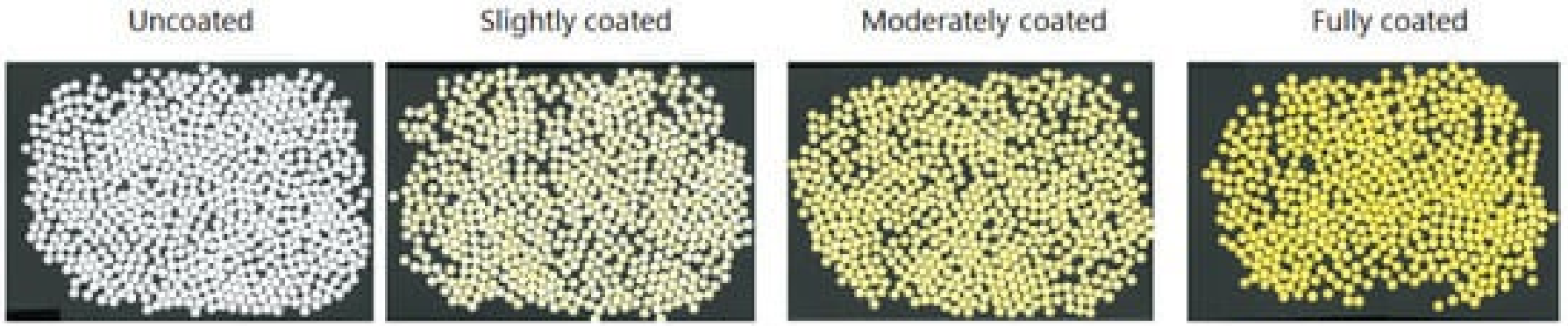


Predicting drug target interaction

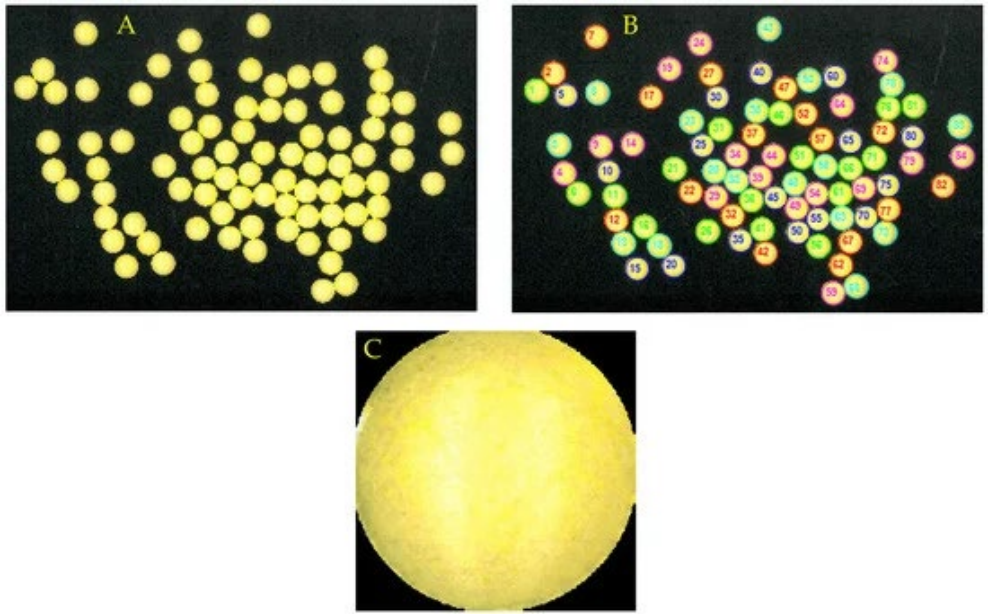


Mesrabadi et. al. (2023) J. Drug–target interaction prediction based on protein features, using wrapper feature selection. *Sci Rep*

AI in manufacturing



Quality assurance of tablet coating quality



Hirschberg et. al. (2020) Image-Based Artificial Intelligence Methods for Product Control of Tablet Coating Quality. *pharmaceutics*

AI in biomedicines

Machine learning can learn subtle and complex patterns in high dimensional spaces connecting measurements with outcomes

There are many tools and models available

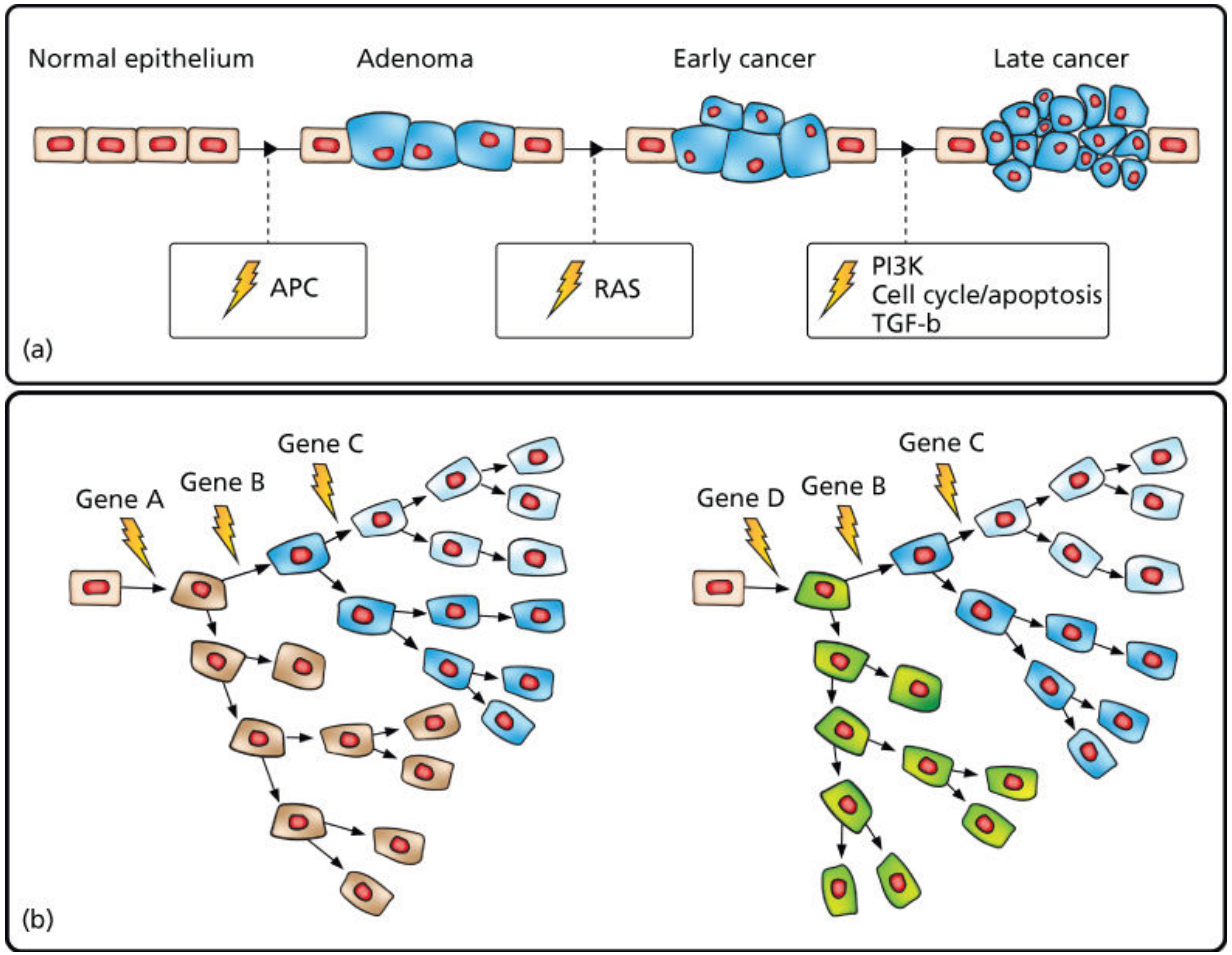
Machine learning has made certain products possible

Neoantigen-based therapies: peptide-MHC/pMHC-T-Cell binding prediction

CRISPR-Cas off target site editing prediction

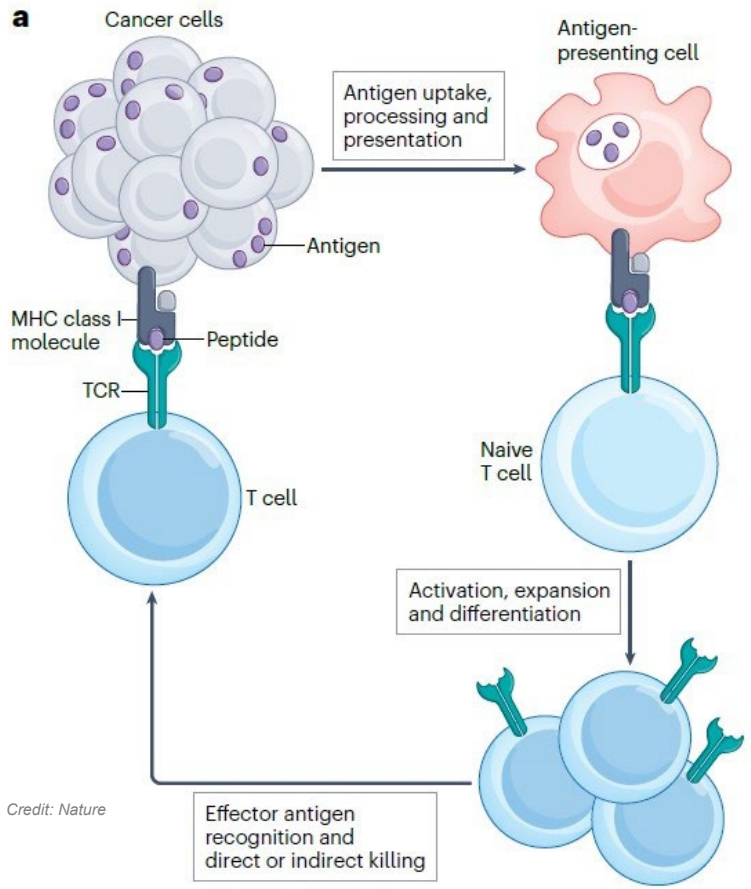
AI IN NEO-ANTIGEN-BASED PRODUCTS

Cancer progression and mutation accumulation

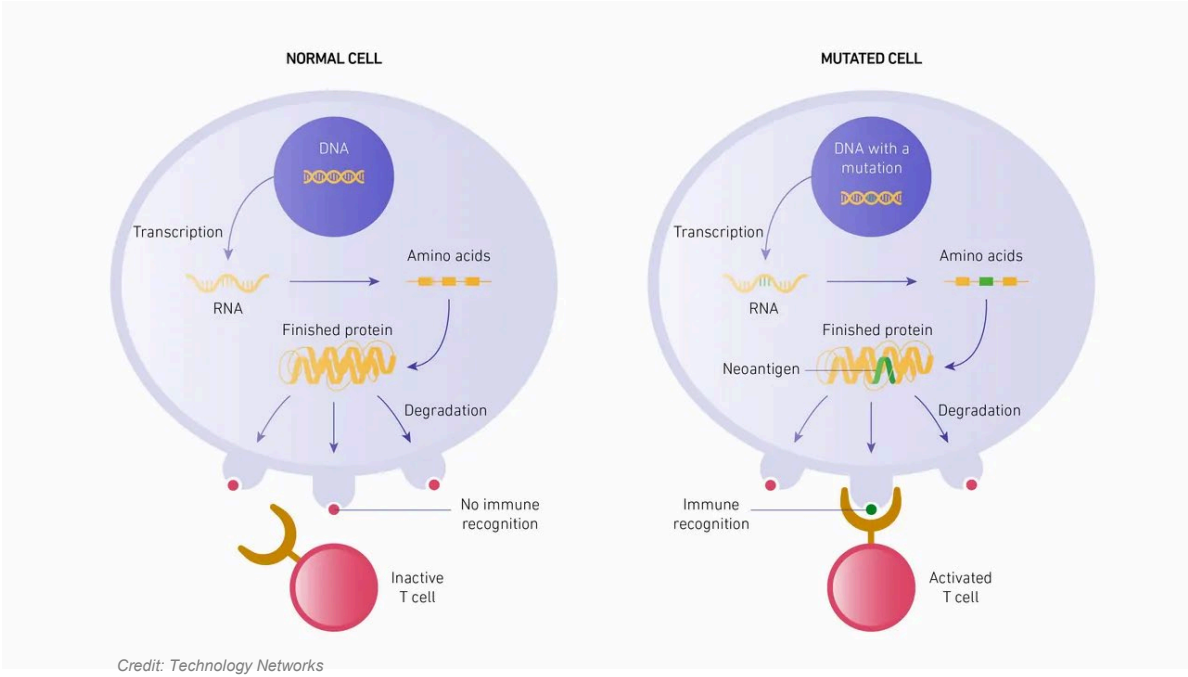


Credit: Oncohema Key

Adaptive immune response to cancer neoantigens



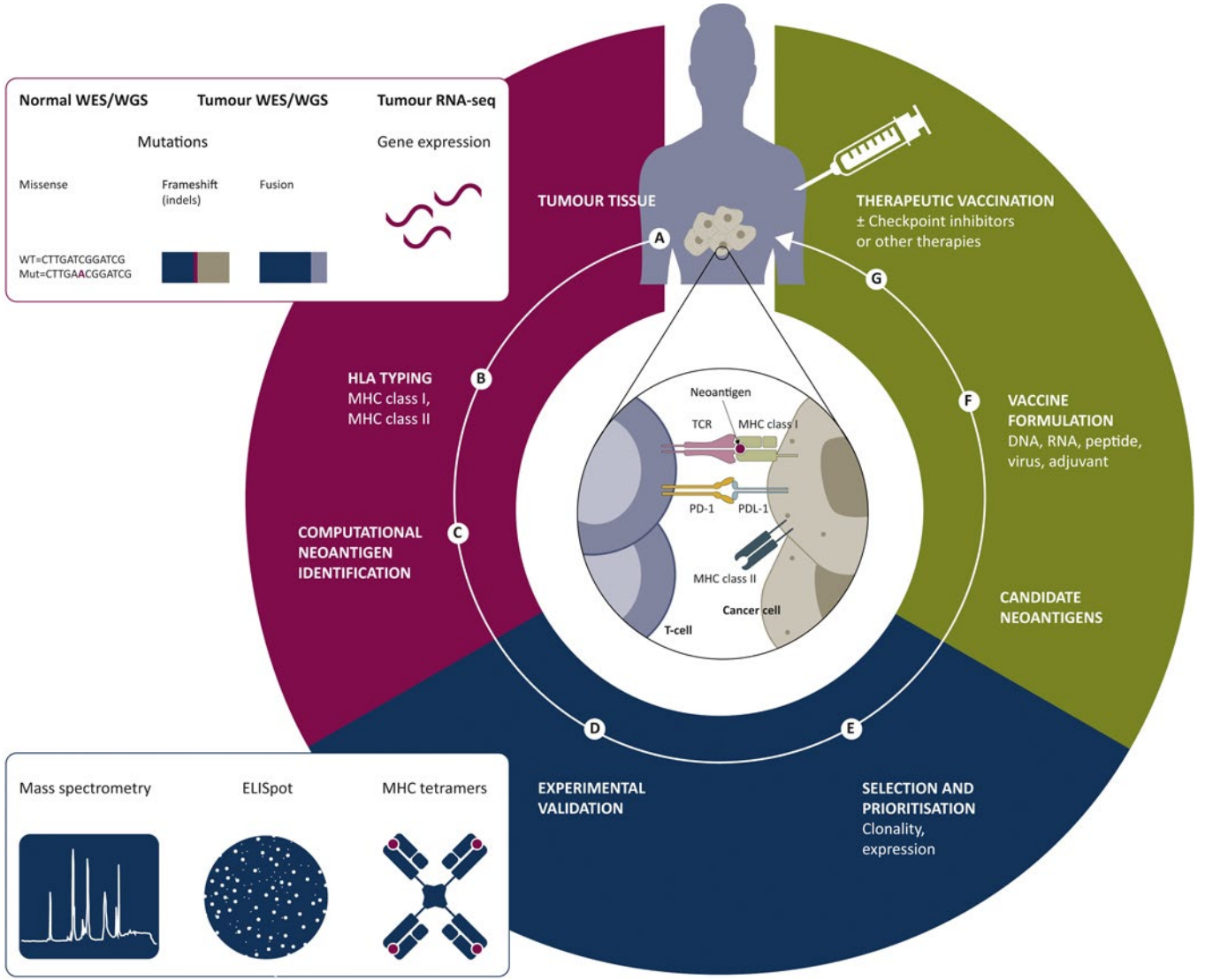
Credit: Nature



Credit: Technology Networks

Neoantigens are newly formed peptides which arise from somatic mutations in cancer cells

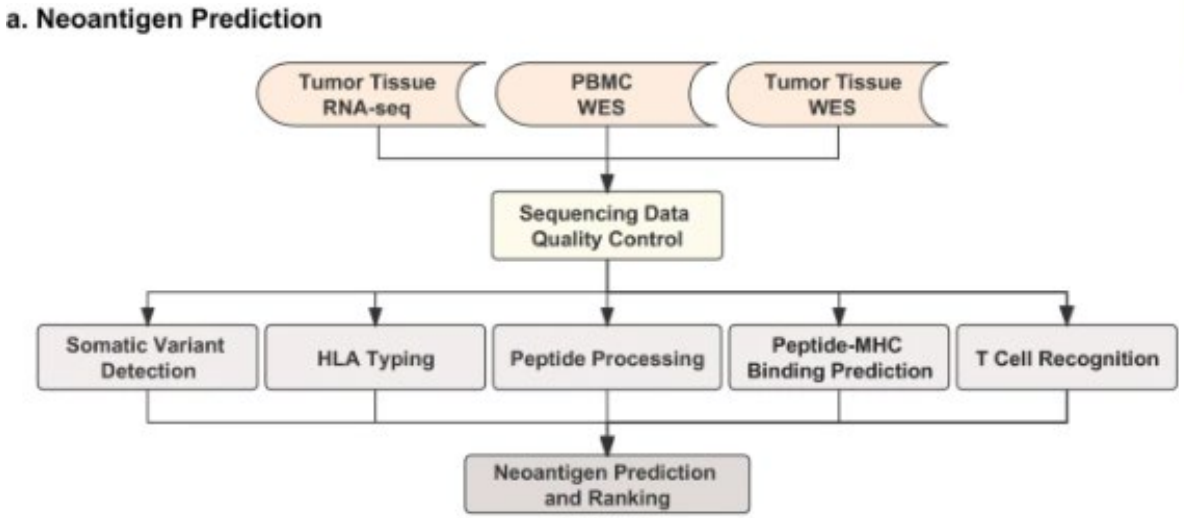
Training the adaptive immune system using neoantigens

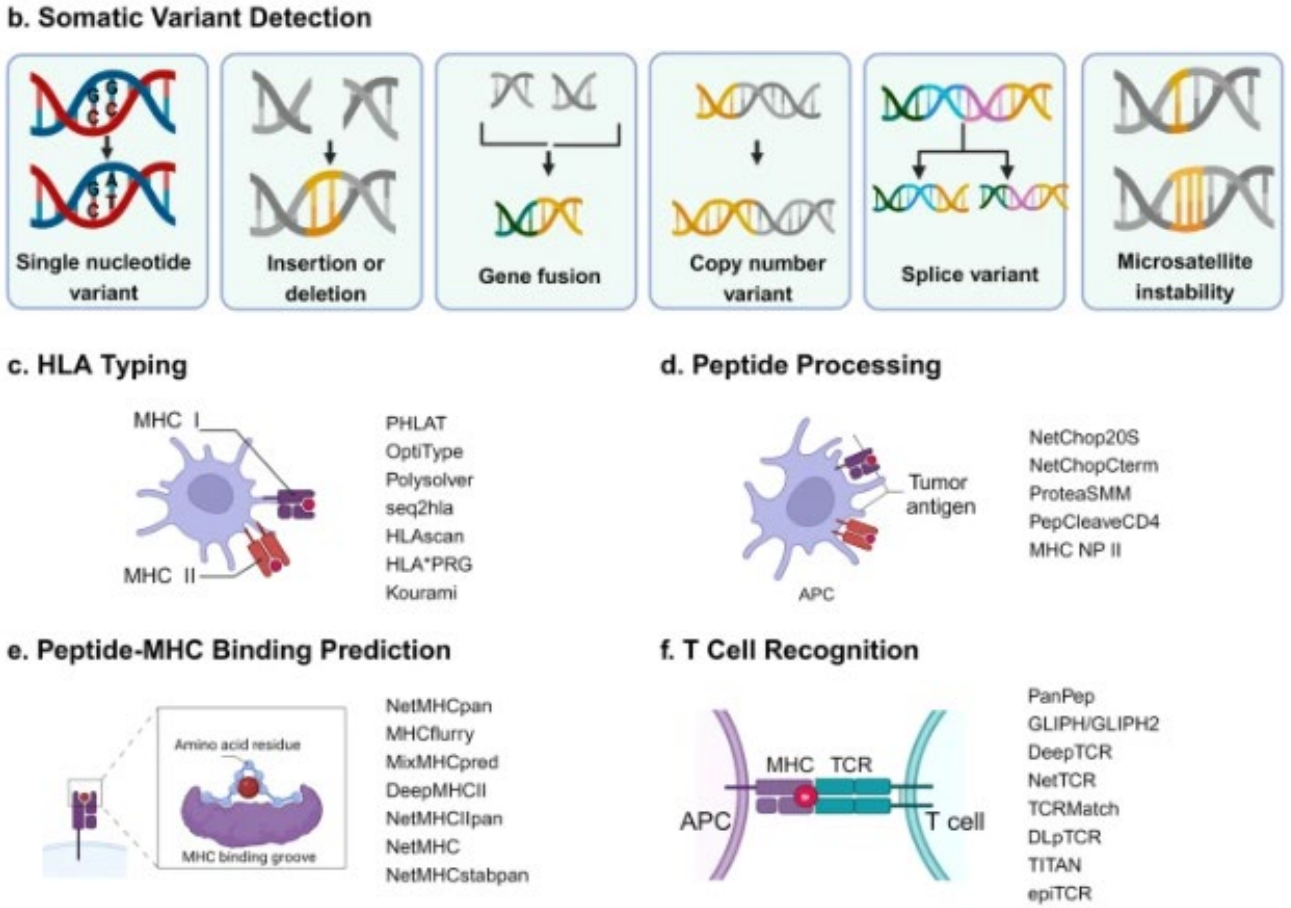


Credit: ScienceDirect

Clinical trials ongoing
Nothing on market

Computational neoantigens identification





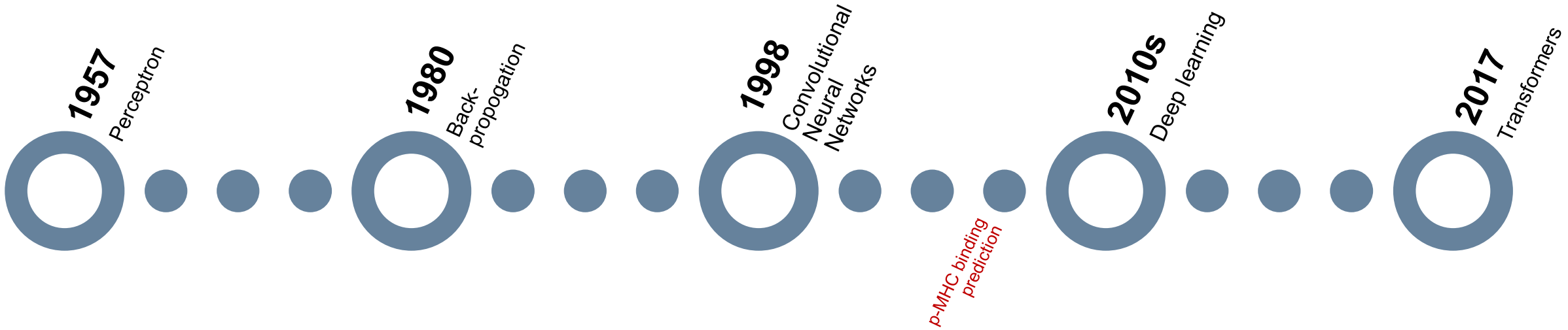
Fan, T. *et al.* (2023) Therapeutic cancer vaccines: advancements, challenges and prospects. *Sig Transduct Target Ther.*

How to predict peptide binding affinity?

?

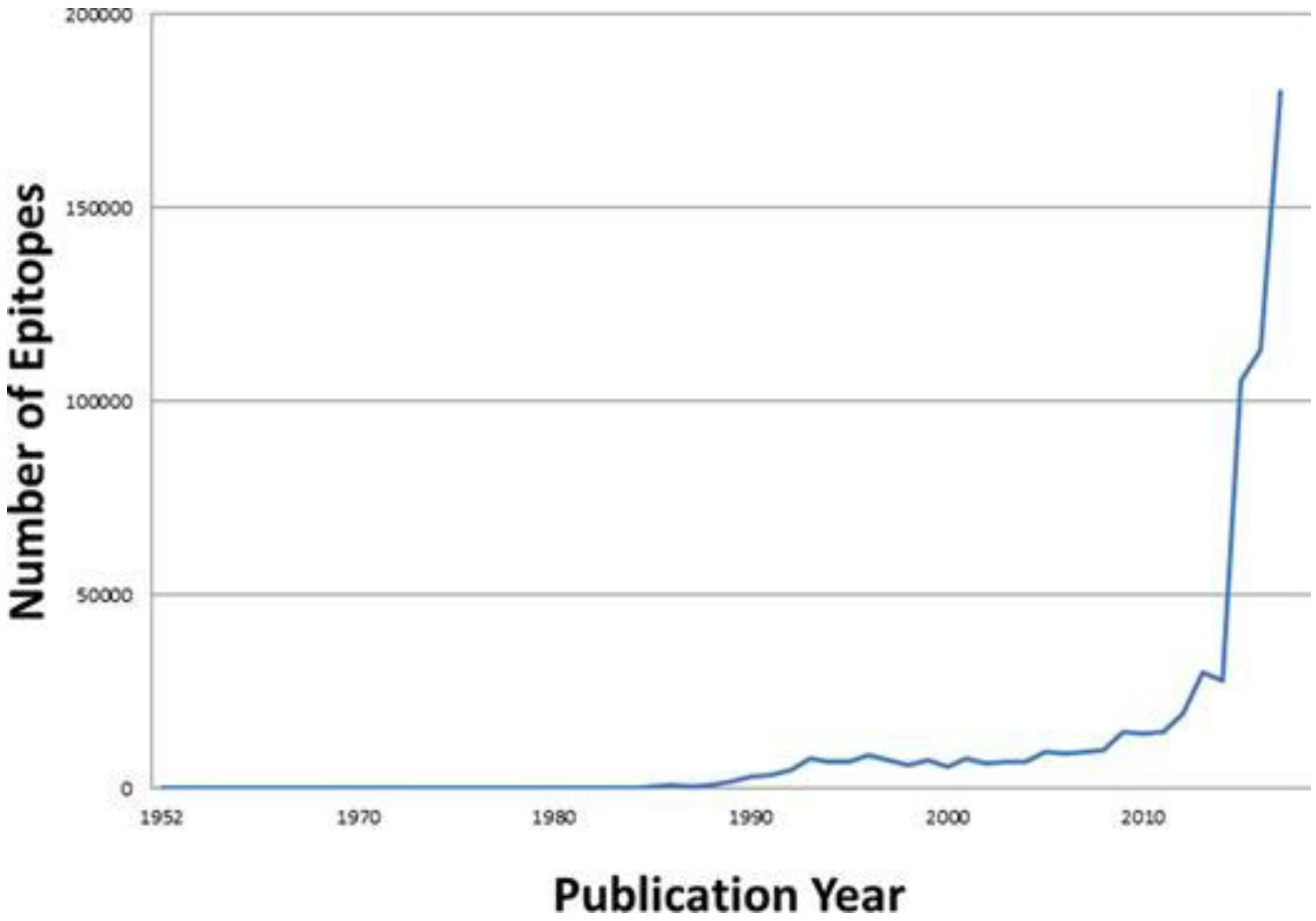
Peptide	Pseudo sequence	Log-affinity
AEFWDVFLS	YFAMYGEKVAHTHVDTLYVRYHYYTWAVLAYTWY	0.0847
ADPVDAVIN	YYAMYGEKVAHTHVDTLYVRYHYYTWAVLAYTWY	0.2890
IRHHVRWAL	YHTEYRNICAKTDVGNLYWTYNFYTWAVLAYEWH	0.4350
YIRRNMINK	YYAMYRNNVAQTDVDTLYIMYRDYTWAVWAYTWY	0.5266
KAGQYVTIW	YDSGYREKYRQADVNKLYLWYDSYTWAEWAYTWY	0.3436
YTAVVPLVS	YTAMYLQNVAQTDANTLYIMYRDYTWAVLAYTWY	0.0014

History of neural networks



Availability of binding affinity data

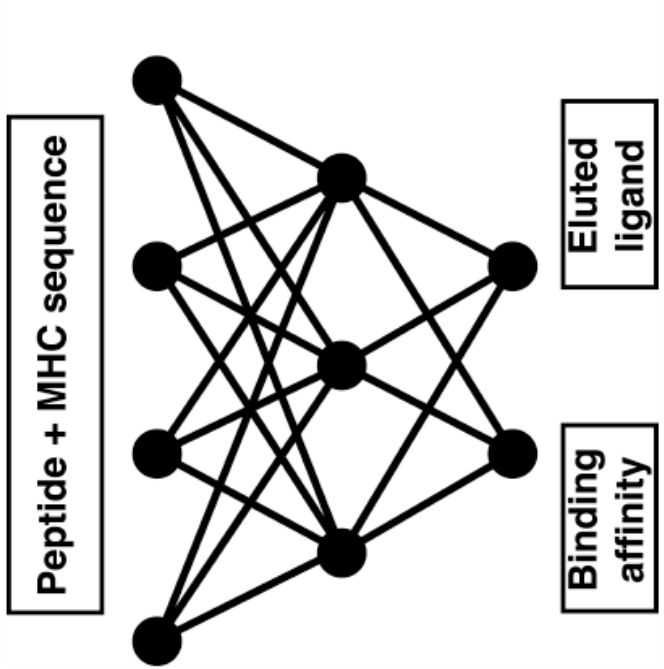
Number of epitopes curated by year in the Immune Epitope Database



Vita R. et. al. (2019) The Immune Epitope Database (IEDB): 2018 update, *Nucleic Acids Research*

How neural networks are applied to peptide-MHC binding prediction

Peptide	Pseudo sequence	Log-affinity
AEFWDVFLS	YFAMYGEKVAHTHVDTLYVRYHYYTWAVLAYTWY	0.0847
ADPVDAVIN	YYAMYGEKVAHTHVDTLYVRYHYYTWAVLAYTWY	0.2890
IRHHVRWAL	YHTEYRNICAKTDVGNLYWTYNFYTWAVLAYEWH	0.4350
YIRRNMINK	YYAMYRNNVAQTDVDTLYIMYRDYTWAVWAYTWY	0.5266
KAGQYVTIW	YDSGYREKYRQADVNKLYLWYDSYTWAEWAYTWY	0.3436
YTAVVPLVS	YTAMYLQNVAQTDANTLYIMYRDYTWAVLAYTWY	0.0014



43 input neurons (one per sequence position)

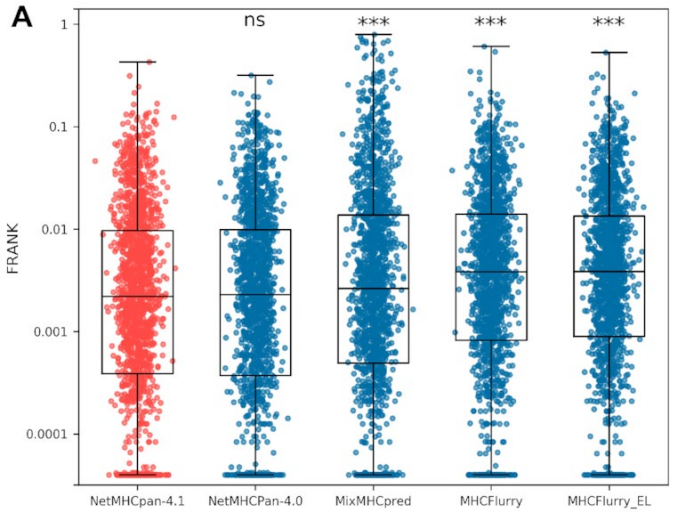
5×10^{50} possible amino acid combinations

Nielsen et. al. (2007). NetMHCpan, a Method for Quantitative Predictions of Peptide Binding to Any HLA-A and –B Locus Protein of Known Sequence. PLoS ONE

Performance

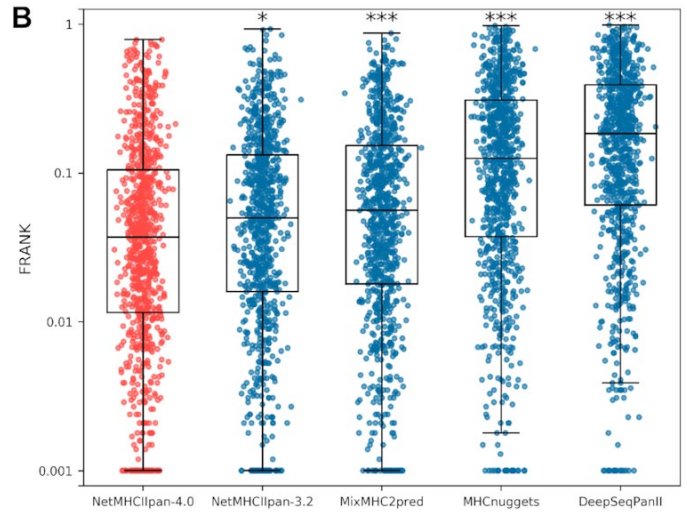
- Binding affinity data:
  - 134281 IC_{50} measurements from IEDB
  - 4 distinct MHC II alleles
- Eluted ligand data:
  - 372639 MHC measurements
  - 74 distinct MHC II alleles
  - Negative peptides sampled from UniProt

CD8+



F-rank: the ratio between the number of peptides with a prediction score higher than the positive peptide and the number of peptides contained within the source protein.

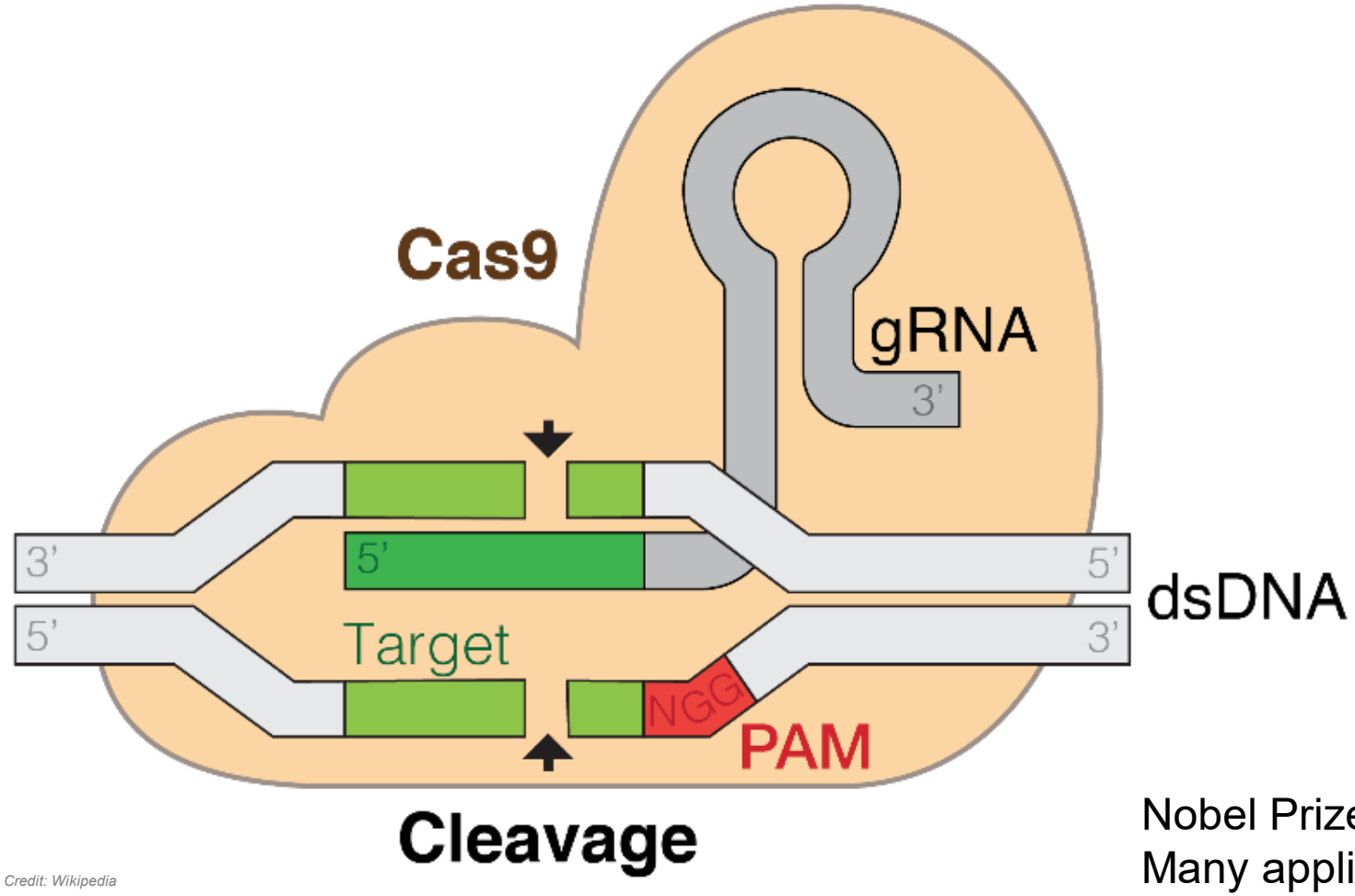
CD4+



Reynisson et. al. (2020), NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data, *NAR*

AI IN CRISPR-CAS-BASED PRODUCTS

CRISPR-Cas9

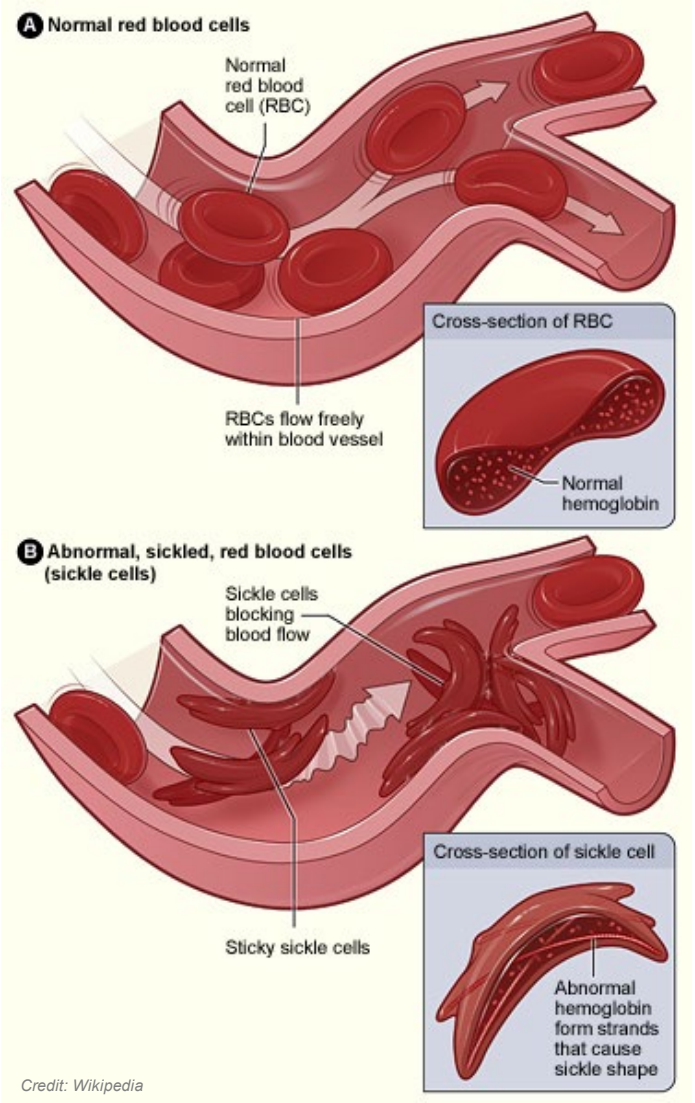


Credit: Wikipedia

Nobel Prize
Many applications
Off-target problems

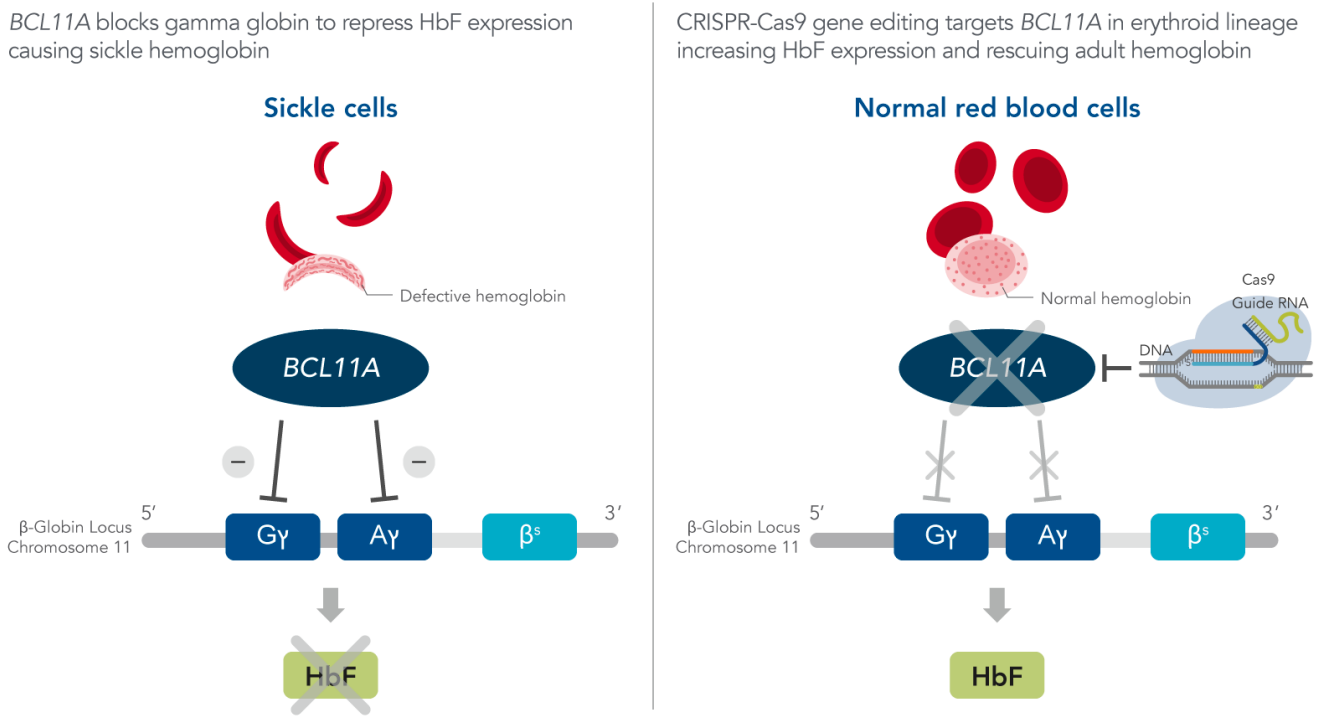
Example indication

Sickle cell anaemia



Credit: Wikipedia

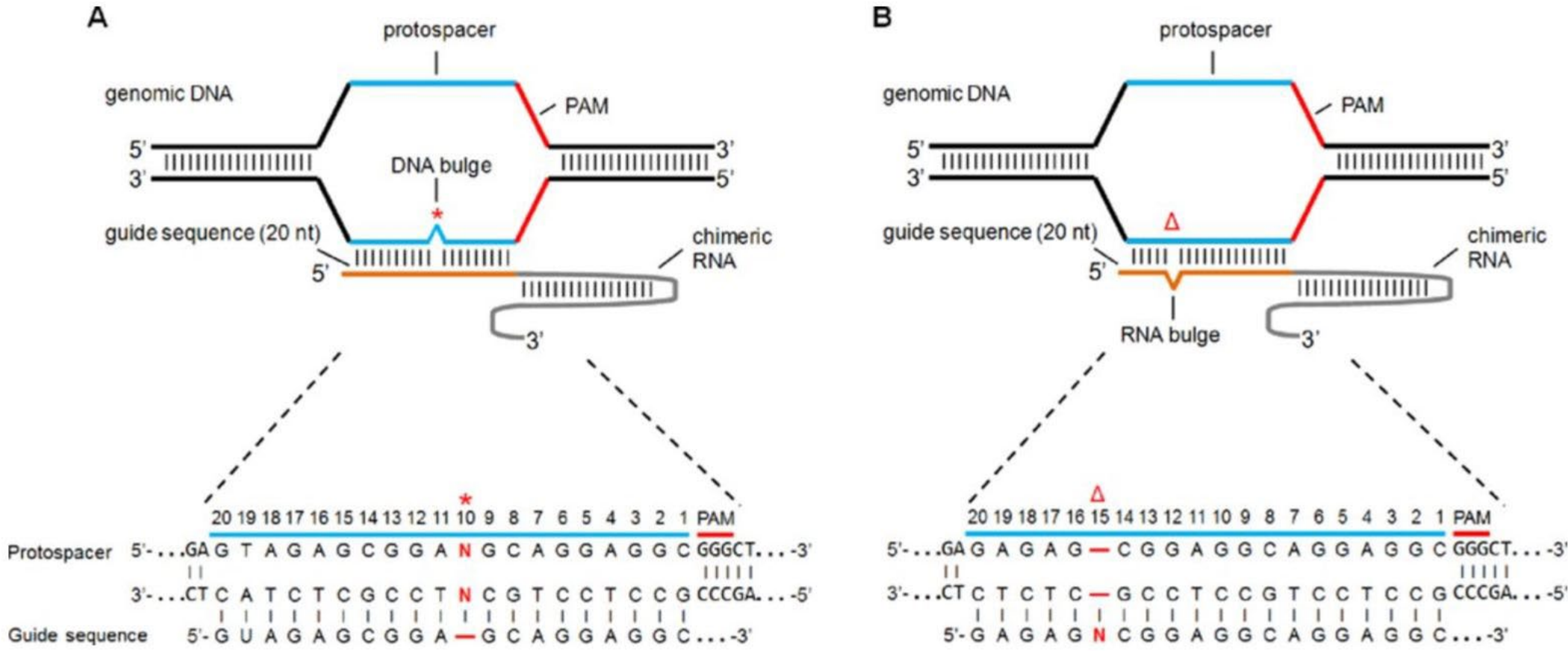
BCL11A KO restores Foetal Hb expression



Credit: Morayo G. Adebiyi, (2024) Tackling sickle cell disease with precise and efficient gene editing. DT

Casgevy
FDA approved December 2023
EMA authorised February 2024

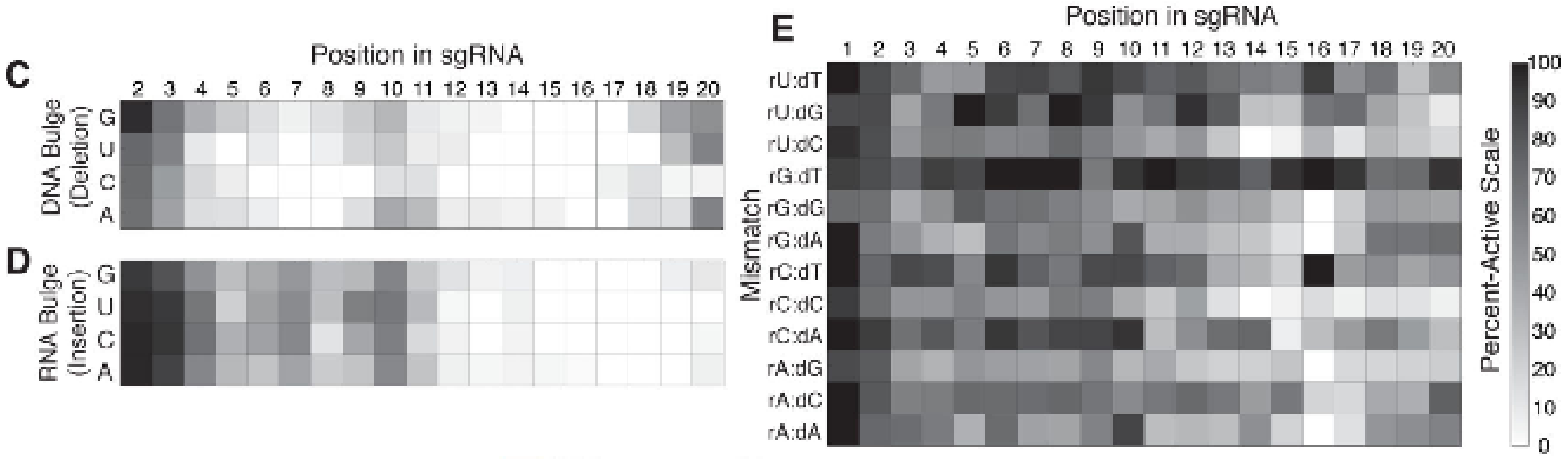
Off-target editing



Lin Y. et. al. (2014) CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences, *Nucleic Acids Research*

Off-target editing site prediction

Mismatch position and identity affect editing activity



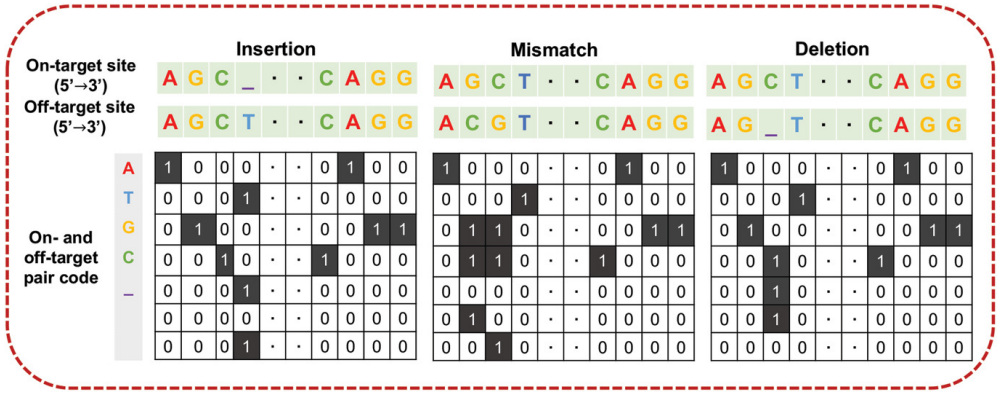
Doench, J. *et al.* (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*

Prediction of off-target site editing

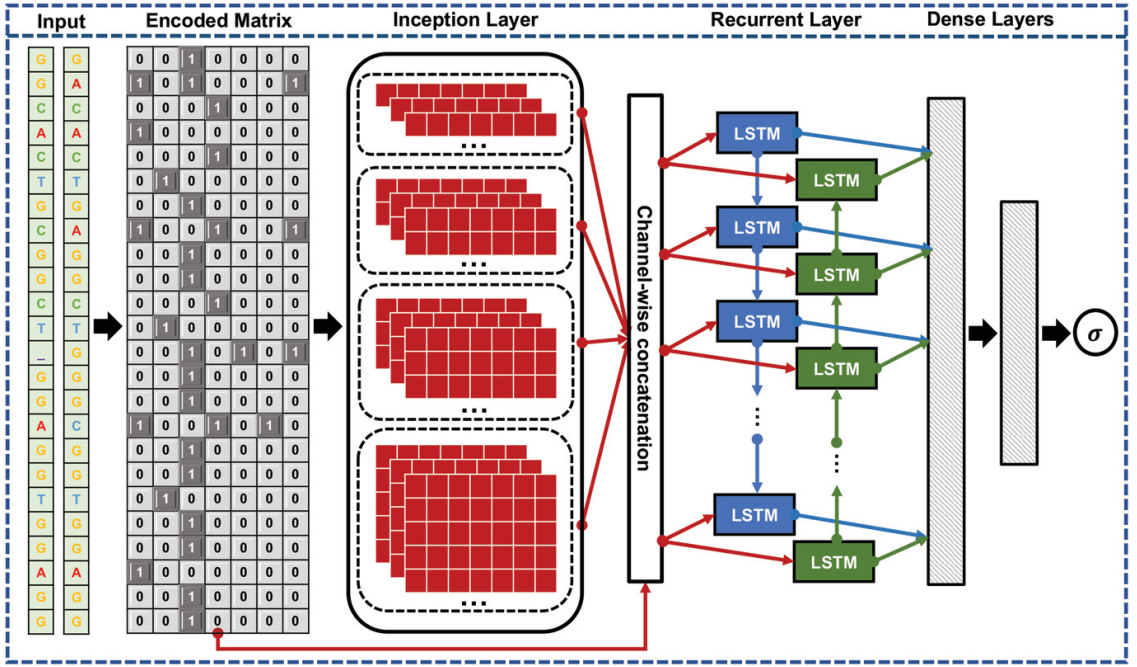
Example input data

Site	Sequence	Mutation (%)
On-target	CTTGCCCCACAGGGCAGTAACGG	78.8
Off-target-1	TCAGCCCCACAGGGCAGTAACGG	85.8
Off-target-3	GCTGCCCCACAGGGCAGCAACGG	4.4

Data encoding



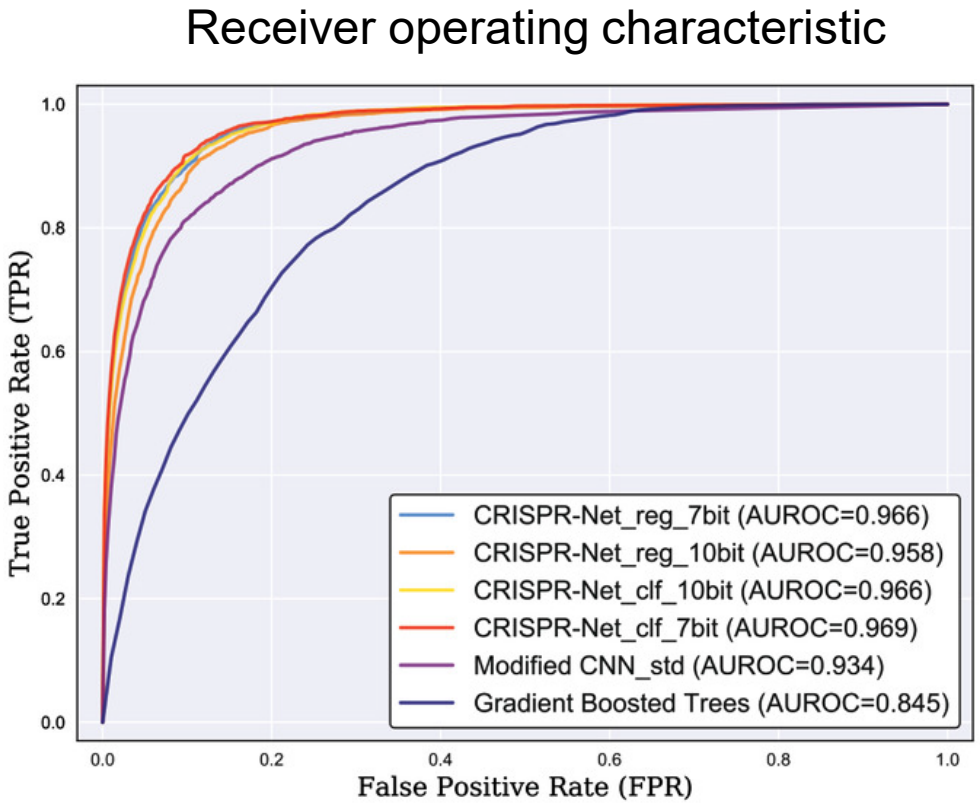
Neural network architecture



Lin et. al. (2020). CRISPR-Net: A Recurrent Convolutional Network Quantifies CRISPR Off-Target Activities with Mismatches and Indels. Advanced Science.

Performance

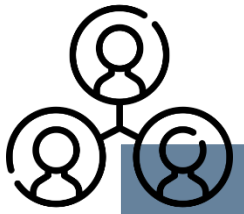
Type / No.	Technique	Total	Validated Off-targets	Guide RNAs	With Indel	Literature
I / 1	CIRCLE-Seq	584 949	7371	10	Yes	Tasi et al.[19]
I / 2	GUIDE-Seq	213 943	60	6	Yes	Listgarten et al.[13]
II / 1	Protein knockout detection	4853	2273	65	No	Doench et al.[11]
II / 2	PCR, Digenome-Seq and HTGTS	10 129	354	19	No	Haeussler et al.[40]
II / 3	SITE-Seq	217 733	3767	9	No	Cameron et al.[18]
II / 4	GUIDE-Seq	294 534	52	9	No	Tasi et al.[16]
II / 5	GUIDE-Seq	95 829	54	5	No	Kleinstiver et al.[44]
II / 6	GUIDE-Seq	383 463	56	22	No	Listgarten et al.[13]



Kim et. al. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nature Methods*
Lin et. al. (2020). CRISPR-Net: A Recurrent Convolutional Network Quantifies CRISPR Off-Target Activities with Mismatches and Indels. *Advanced Science.*

USING AI

Practically integrating AI into your product

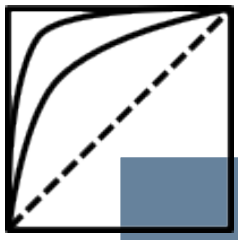


Skilled team

- Data cleaners
- Data scientists
- Bioinformaticians
- Developers
- Domain experts
- System admins
- Scrum masters
- UI/UX experts
- Product owners

Infrastructure

- Cloud
- Servers
- Cluster
- Linux
- Containerisation
- Continuous integration



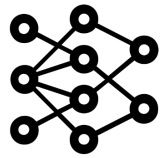
Validation

- Technical
  - Unit testing
  - Integration testing
  - Regression tests
- Analytical
  - Cross validation
  - Independent data
- Clinical

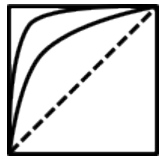
What to consider when using AI in the development of medicinal products

What risks are inherent in each step?

- Risk of false predictions
- Risk of tools used



Should we develop our own machine learning models?



How well does the step perform?

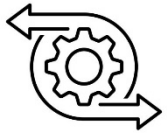
- Analytical validation as part of risk assessment

How are risks mitigated?

- Accept, avoid, transfer, reduce
- Verification, filtering

What dataset should be used?

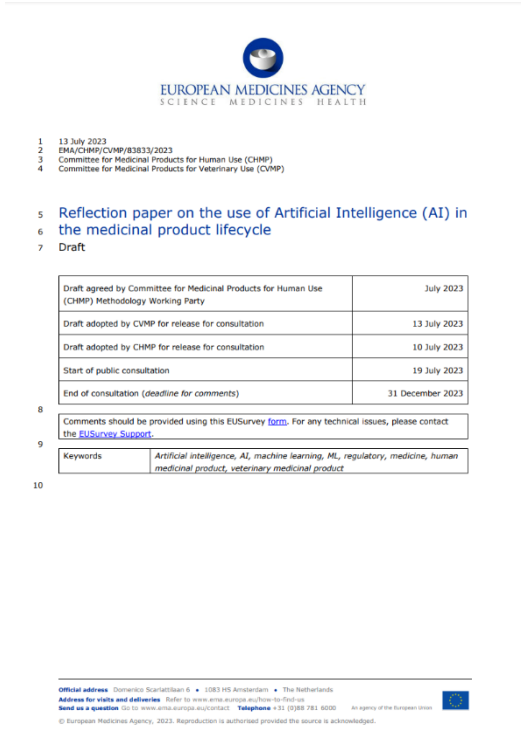
- Published, purchased, in-house, synthetic



How is change managed?

- Major/minor changes
- Reporting

Regulatory guidelines


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*

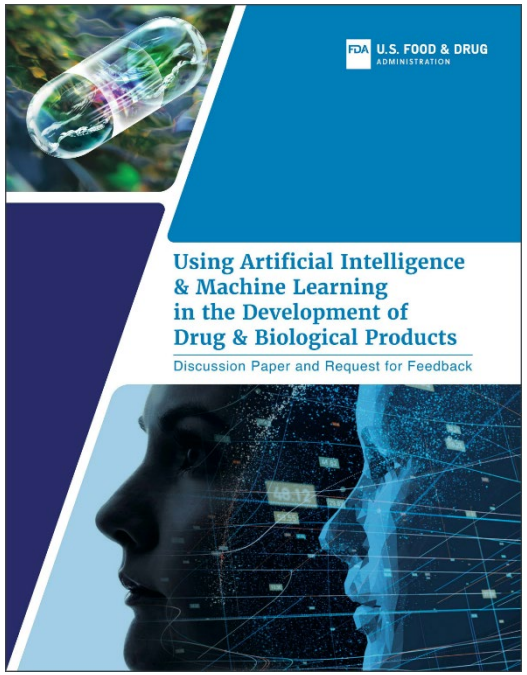

EU AI Act

Proposal for a

Regulation of the European Parliament and of the Council Laying Down Harmonsed Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts

2021/0106 (COD)

European Commission



EMA: https://www.ema.europa.eu/en/news/reflection-paper-use-artificial-intelligence-lifecycle-medicines
EU: https://artificialintelligenceact.eu/
WHO: https://www.who.int/publications/i/item/9789240078871
FDA: https://www.fda.gov/science-research/science-and-research-special-topics/artificial-intelligence-and-machine-learning-aiml-drug-development

OUTLOOK

Outlook

AI is becoming an integral part of the medicinal product life cycle.

The discussion of how AI will be regulated is ongoing (see slide 36)

Two examples where AI could be used directly in design of biomedicine

Would love to hear about more

Acknowledgements

Cellular aspects of Pathogen-Host-Interactions research group

Renate König

Leona Enke

Zsófia Nacsa

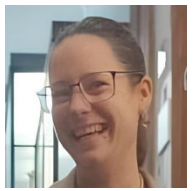
AI and Big Data working group

Martin Machyna

Liam Childs

Markus Braun

Farnaz Zeidi



Maike Hermann

Roman Christof

Alla Bulashevska

HZG 2 Blood products and gene therapies

Silvia Vogl

Z5 Information Technology

Johannes Weydt

THANK YOU VERY MUCH

FOR YOUR ATTENTION