The current and emerging use of AI in safety surveillance post-marketing

Andrew Bate

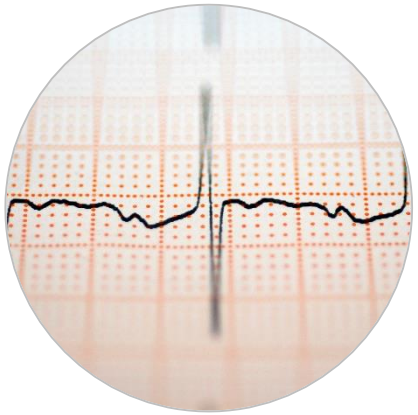
AIRIS February 27th, Seoul, Korea

Disclosures

- I am a full-time employee of GSK (hold stock and stock options)
- Previously at Pfizer, prior to that headed Research at Uppsala Monitoring Centre
- Honorary Associate Professor of Epidemiology, LSHTM, UK
- Former Visiting Full Professor, Information Systems, Brunel University
- Former Adjunct Associate Professor of Clinical Pharmacology, NYU, USA

Spontaneous reporting as a tool for post marketing surveillance

Analysis of spontaneous reports of suspected adverse drug reactions (ADRs) is a valuable tool in the **detection of previously unknown** drug adverse reactions



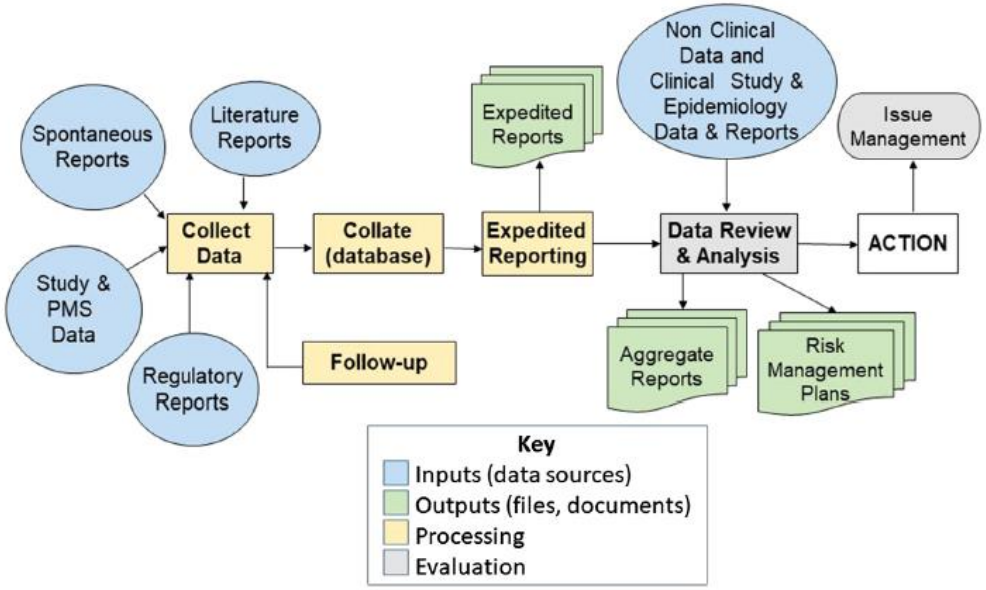
Reports of adverse events (AEs) associated with a drug are **not necessarily true ADRs**, that is, they may be temporally associated with a drug but not caused by the drug

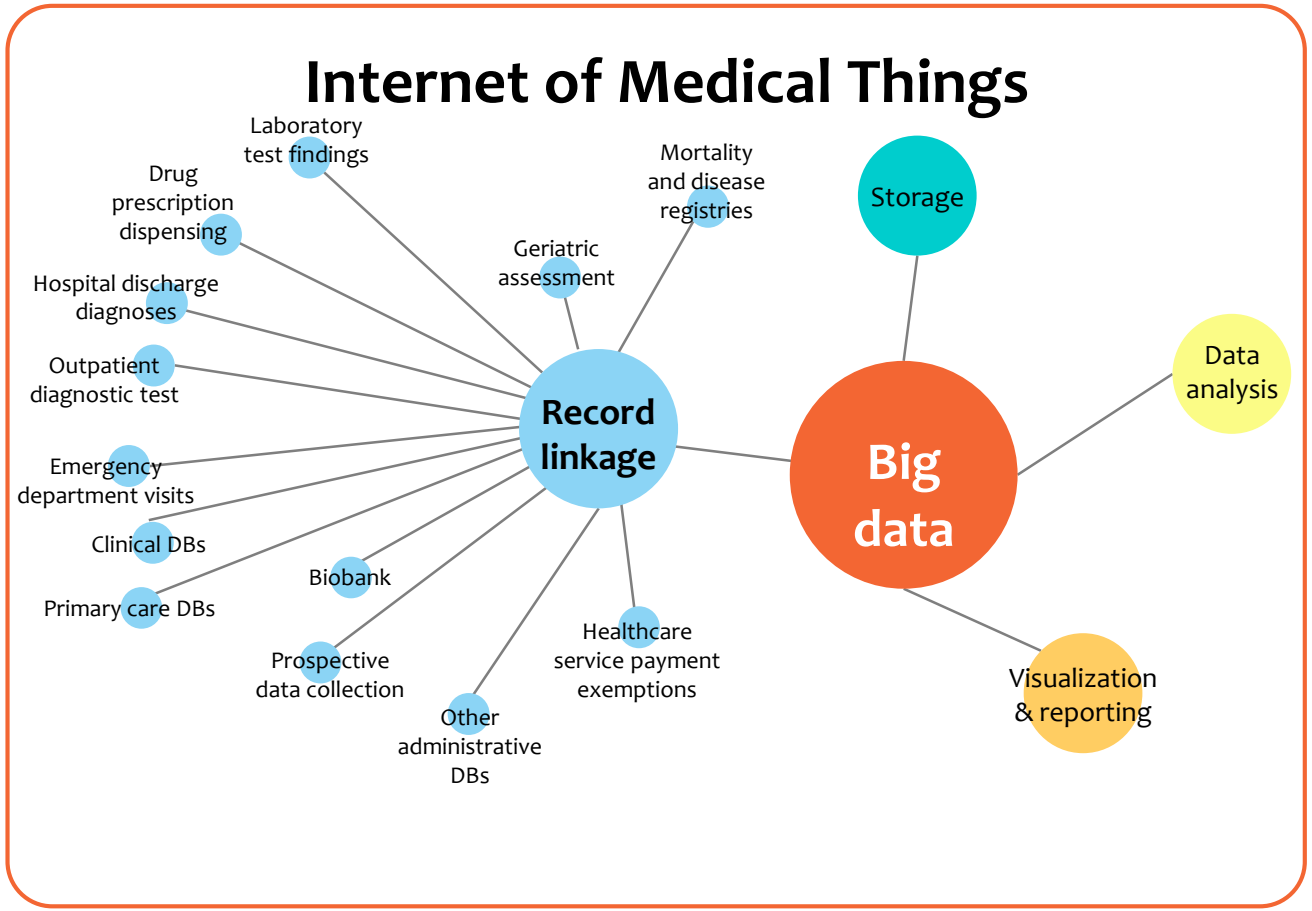
Hypothesis generation of new possible adverse reactions from such data is referred to as **signal detection**

Bate, A. and Evans, S.J.W. (2009), Quantitative signal detection using spontaneous ADR reporting. Pharmacoepidem. Drug Safe., 18: 427-436.

GSK

Highly Complex Multiple Data Stream Pharmacovigilance Lifecycle – Ripe for Widespread Automation





- Lewis, DJ and McCallum, JF. Utilizing advanced technologies to augment pharmacovigilance systems: challenges and opportunities. *Ther Innov Regul Sci.* 2020;54:888-899.

DB, database

Adapted from Trifirò G, et al. From big data to smart data for pharmacovigilance: the role of healthcare databases and other emerging sources. *Drug saf.* 2018;41:143-149.

The Potential Benefits Of Implementing Intelligent Automation Technologies In PV

Intelligent Automation Benefits

Optimize Existing Processes
reduce cycle times, manage volume

Improve Quality
meet then exceed human performance

Redirect Resources
to value added activities

Gain Insight
find interesting "index cases" first

Unlock New Data Sources
real world & unstructured

Applicable PV Areas

Individual Case Safety Report (ICSR) Process
- *Acquiring & Integrating Safety-Relevant Data*
- *Case Intake, Processing, & Reporting*

Periodic Reporting
- *Aggregate ICSR Data*

Signal Detection & Management
- *Analytics & Decision Support*

Risk Management
- *Risk Decision-Making*
- *Risk Communication & Education*

Outcomes

Improved Insights and Access to Better Information, Leading to Enhanced Patient Safety and More Informed Healthcare Decisions

From: TransCelerate's Intelligent Automation Opportunities in Pharmacovigilance

GSK

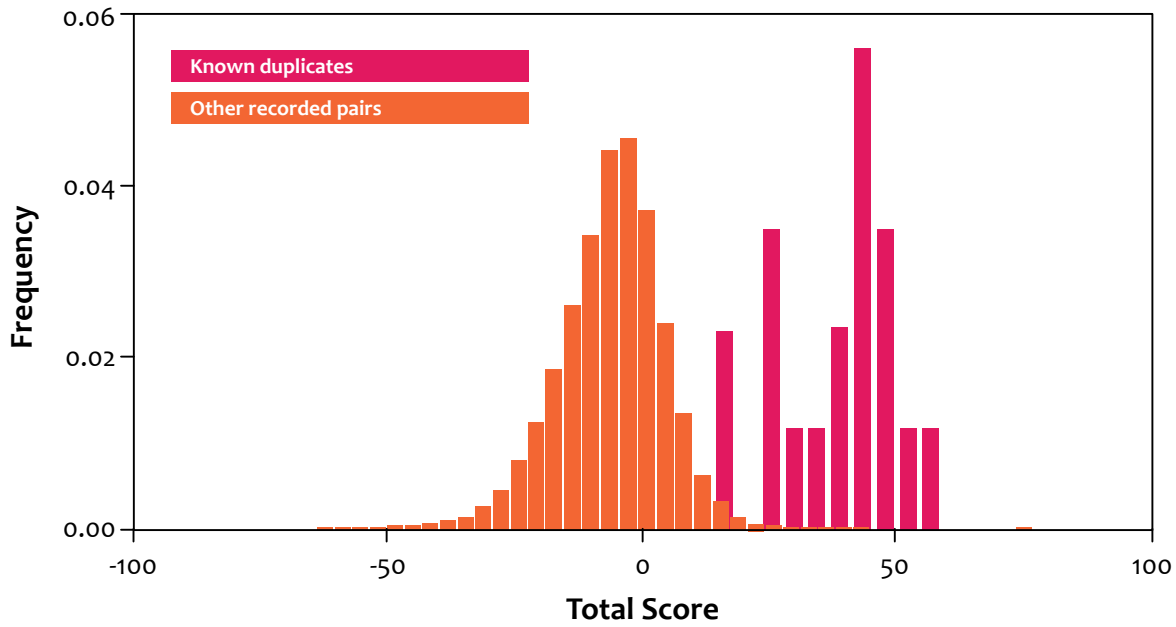
Routinely used ML for duplicate detection of safety reports

Published: 07 February 2007

Duplicate detection in adverse drug reaction surveillance

G. Niklas Norén, Roland Orre, Andrew Bate & I. Ralph Edwards

Data Mining and Knowledge Discovery **14**, 305–328 (2007) | Cite this article

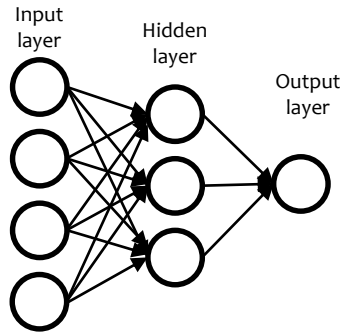


Age	Gender	Country	Drug substances	ADR terms	Onset date	Outcome	Score
51	F	NOR	6 matched, 1 unmatched	3 unmatched	2004-04-30	?	+76.97
50	F	NOR			2004-04-20	?	

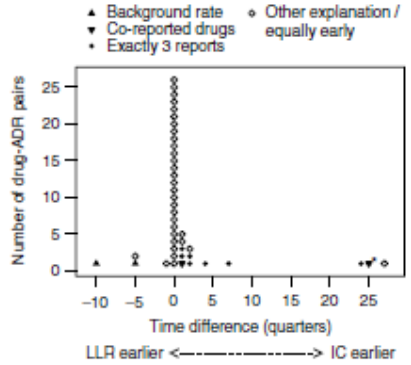
- This case pair: Near-matches on age and date, no matching ADR terms but 6 matching drug substances (not commonly co-prescribed) ADR terms are semantically close (not shown). Note also Interested in sets of 3+ similar reports

GSK

ML Method application – not new! How useful has it been in practice?



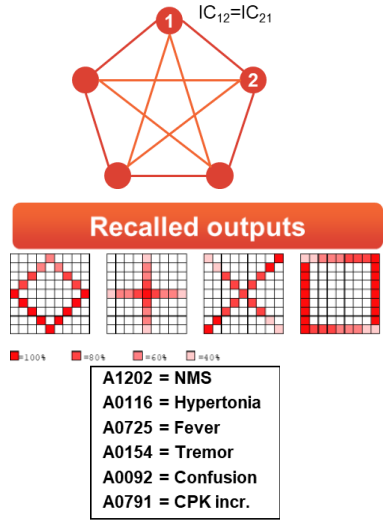
Alvager T et al. The use of artificial neural networks in biomedical technologies: an introduction. *Biomed Instrum Technol.* 1994;28:315-322.



- Lasso shrinkage regression not clearly better than disproportionality

Caster O, et al. Large-Scale Regression-Based Pattern Discovery: The Example of Screening the WHO Global Drug Safety Database. *Stat Anal Data Min.* 2010;3:197-208.

Cij	i	A1202	A0116	A0725	A0154	A0092	A0791	A0163	A0091	A0093	A0224	A0151	A0210	A0043	A0280	A0576	A0156	A0507
j	Ci/Cj	723	585	517	357	348	270	217	174	174	145	143	125	108	108	92	66	60
A1202	723	-	109	171	**23**	**29**	126	**17**	20	**11**	39	24	27	22	**8**	43	6	**3**
A0116	585	109	-	121	67	43	88	26	20	13	16	40	24	26	19	17	8	6
A0725	517	171	121	-	33	38	109	18	25	14	47	30	43	35	9	38	**3**	4
A0154	357	**23**	67	33	-	24	**9**	**8**	**6**	11	11	12	8	20	8	**3**	5	**1**
A0092	348	**29**	43	38	24	-	25	39	**7**	9	14	10	8	11	5	10	16	**2**
A0791	270	126	88	109	9	25	-	7	13	**5**	25	14	19	11	5	47	5	5
A0163	217	**17**	26	18	**8**	39	7	-	6	6	5	10	5	5	4	**2**	3	2
A0091	174	20	20	25	**6**	**7**	13	6	-	19	9	**2**	5	6	**2**	**1**	5	6
A0093	174	**11**	13	14	11	9	**5**	6	19	-	3	8	8	**1**	3	**1**	**1**	7
A0224	145	39	16	47	11	14	25	5	9	3	-	6	29	18	2	7	**1**	**1**
A0151	143	24	40	30	12	10	14	10	**2**	8	6	-	**1**	5	4	3	2	5
A0210	125	27	24	43	8	8	19	5	5	8	29	**1**	-	4	2	6	3	1
A0043	108	22	26	35	20	11	11	5	6	**1**	18	5	4	-	5	3	2	1
A0280	108	**8**	19	9	8	5	5	4	**2**	3	2	4	2	5	-	1	1	2
A0576	92	43	17	38	**3**	10	47	**2**	**1**	**1**	7	3	6	3	**1**	-	2	1
A0156	66	6	8	**3**	5	16	5	3	5	**1**	**1**	2	3	2	1	2	-	2
A0507	60	**3**	6	4	**1**	**2**	5	2	6	7	**1**	5	1	1	2	1	2	-



Orre R, et al. A bayesian recurrent neural network for unsupervised pattern recognition in large incomplete data sets. *Int J Neural Syst.* 2005;15:207-222.

Good machine learning practice - a systematic review

- If consider criteria: 1. Large datasets, 2. The use of pretrained models when appropriate, 3. Method novelty, and 4. Reproducibility
  - Reviewers' subjective evaluation found that 42 (10%) studies were reflective of modern best practices in ML/deep learning
- Vast majority (73%) used 'off-the-shelf' methods with little to no problem-specific adaptation or domain knowledge
- Similarly, 92% trained a model 'from scratch', ie only 8% leveraged a pretrained model in some capacity, and only 18% explicitly used some kind of external information or data
- 63% percent of the studies used data that were publicly available (but this included use of social media), while 7% had code that was publicly accessible at some point in time
- Of note: 10% of studies reported no explicit sample size at all

- Ref Kompa B et al. 2022 Artificial Intelligence Based on Machine Learning in Pharmacovigilance: A Scoping Review. Drug Safety. 45 (5), 477-491

GSK

Automation is standard at GSK – three specific data ingestion examples

Results – Data quality reviewer (DQR) automation

DQR Automation workflow

- Daily work allocation report
- Locate cases for processing
  - Log into safety database
  - Search for each provided case id (one at a time)
- Extract current case data
  - Open each case, extract relevant fields, and update output report
- Finalize output
  - Highlight output fields with discrepancies (based on defined business rules)
  - Upload quality reviewed output to original case in safety database
- Review output
  - Safety case processor reviews all output

Human activity

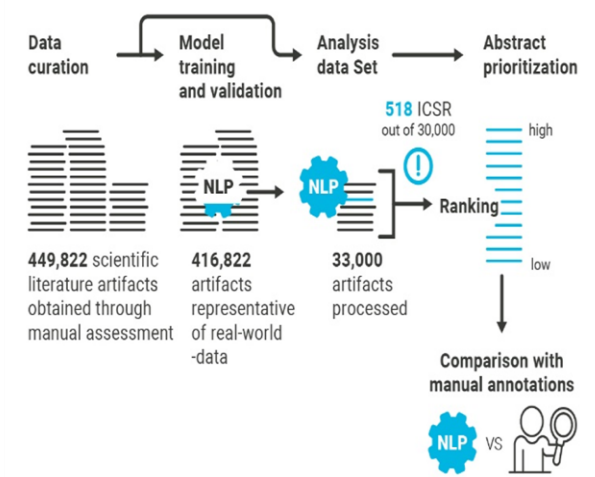
Automated activity

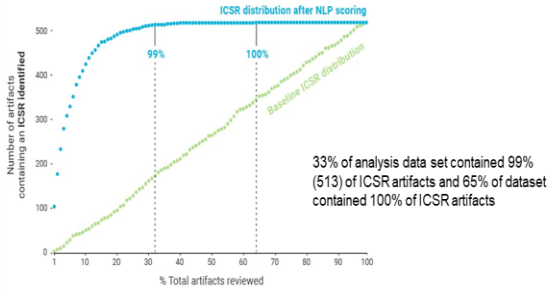
DQR Automation results, single week in March 2020	
Cases processed	458
Time saved per case	602 seconds
Total time savings	~76 hours

DQR Automation results, March to September 2020	
Cases processed	30,702
Time saved per case	602 seconds
Total time savings	~5,134 hours

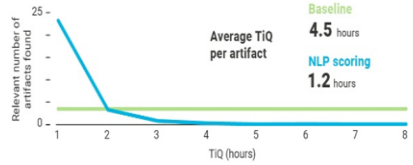
Kassekert R, et al. Automation in routine use for data collection and processing for scalable faster RWE generation. *Value Health*. 2022. In Press.

Could NLP be used to prioritize the artifacts for review?

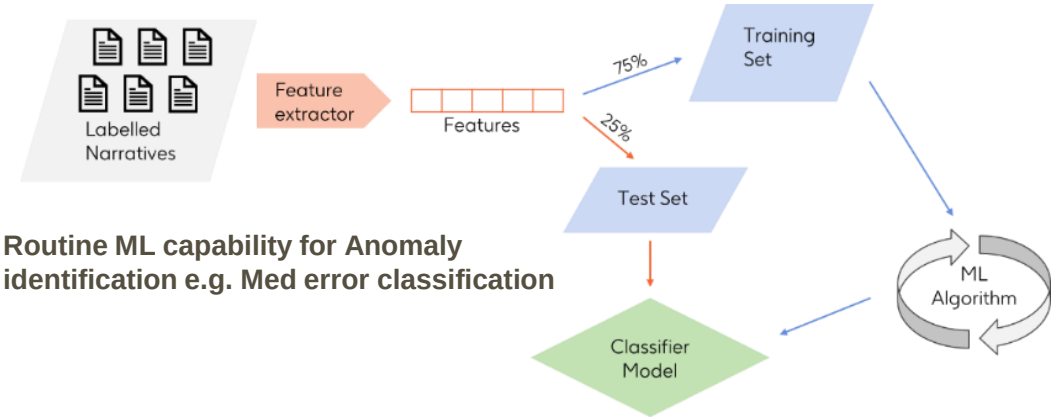




NLP scoring had **a lower Time in Queue** (TiQ)



Glaser et al, 2021, International Conference of Pharmacoepidemiology (poster presentation)



Routine ML capability for Anomaly identification e.g. Med error classification

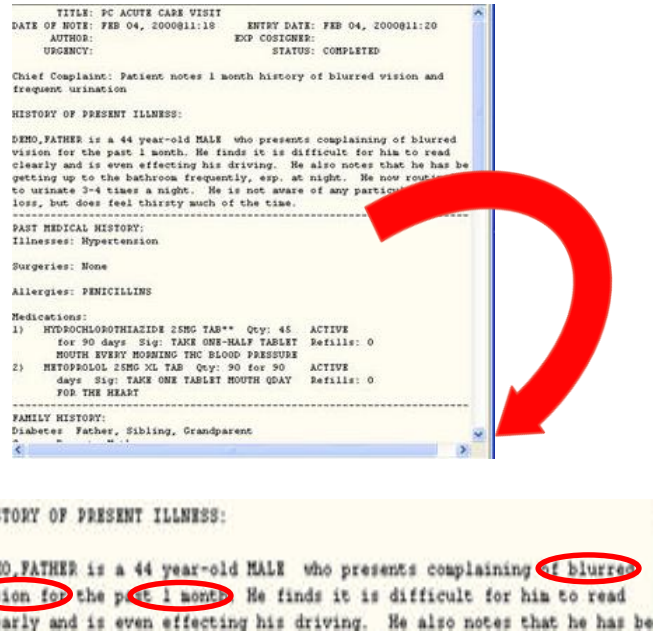
Harnessing the value of unstructured RWD through NLP

Structured (coded)

	Gender (M/F)	Age	Weight (lbs.)	Height (in.)	Smoking (1=No, 2=Yes)	Race
Patient #1	M	59	175	69	1	White
Patient #2	F	67	140	62	2	Black
Patient #3	F	73	155	59	1	Asian
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Patient #75	M	48	90	72	1	White

Demographics. diagnoses, procedures, Rx, lab orders &/or results, billing, operations data

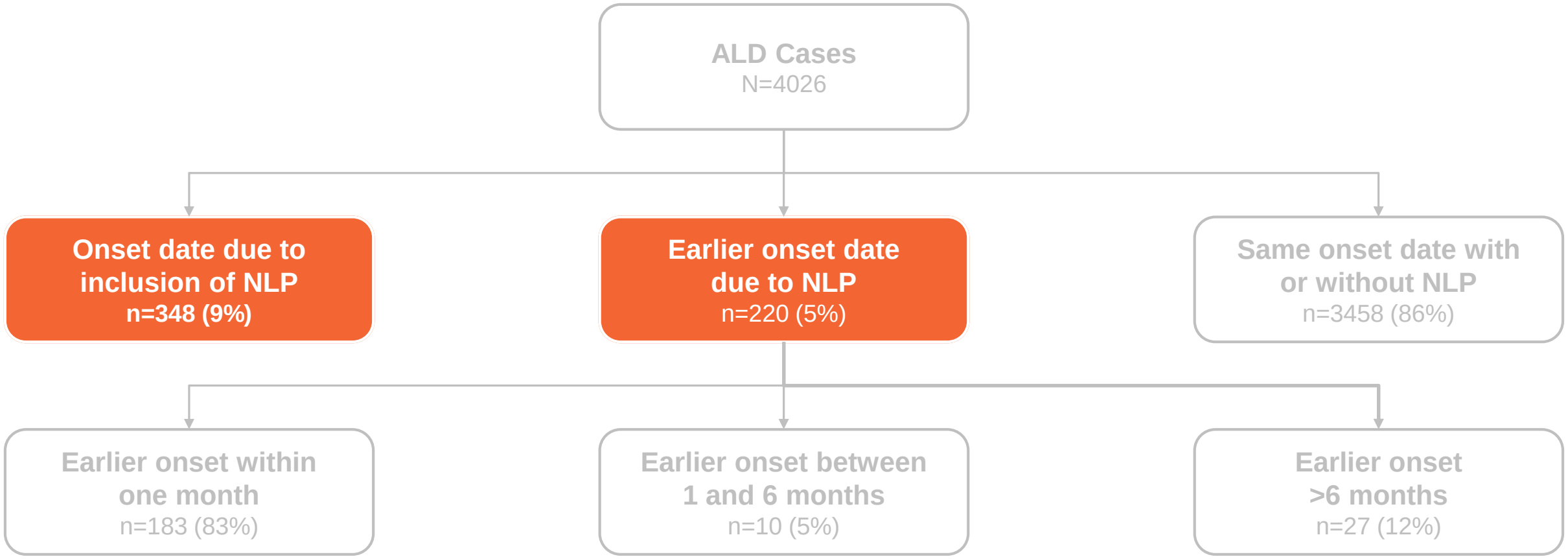
Unstructured (free text)



For an applied example (acute liver injury) see Walker A *et al*

NLP, natural language processing; RWD, real-world data.
Walker AM et al. Int J Med Inform. 2016 Feb;86:62-70.

NLP data contributed to defining ALD onset dates and getting earlier ALD onset dates through gleaned insights from EHR unstructured clinical notes



ALD, acute liver disease; NLP, natural language processing

Walker AM et al. Int J Med Inform. 2016 Feb;86:62-70.

Multiple data stream strategy – still evolving across the field – examples

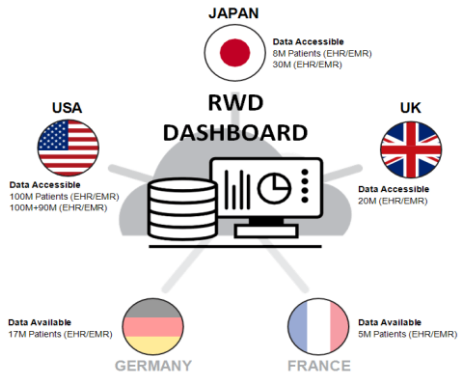
Data

Data ingestion	Data storage	Data cleaning	Data curation	Data analysis and enrichment
Acquire and merge data (internal, external aggregators and web scraping)	1. Gather source docs 2. Organize meta-data (e.g., dates, titles) 3. Add structure for analysis 4. Import into database	1. Filter for spam 2. Remove duplicates 3. Keyword filters for medically relevant posts 4. Apply PII removal	1. Crowdsourcing 2. Expert reviewers 3. Insight Explorer tool 4. ML/AI for protocol-specific questions	1. Linked posts 2. Utility scoring 3. Sentiment analysis 4. NLP Classifiers (e.g., ADR detection) 5. Topic/trend analysis (k-means clustering)

Data oversight and governance

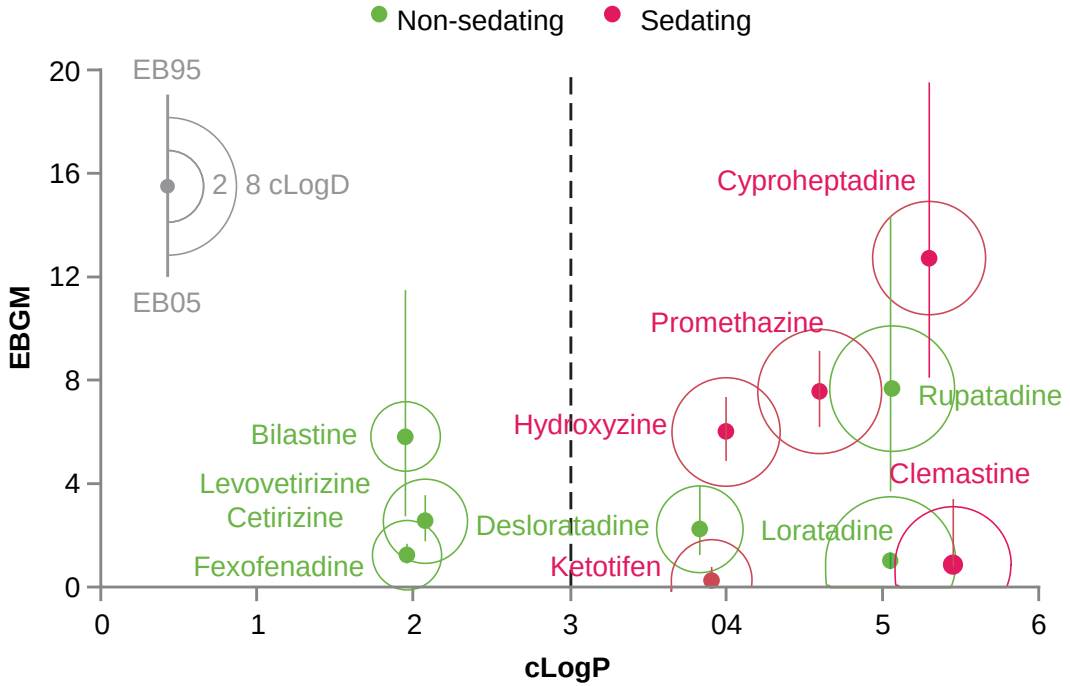
- Data set agnostic social media data platform

Powell G et al. Engaging patients via online healthcare fora: three pharmacovigilance use cases. *Front Pharmacol*. 2022;13:doi: 10.3389/fphar.2022.901355



Painter JL et al. Leveraging data pathways for next generation safety monitoring of medicines and vaccines. 2022 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE. IEEE CPS Proceedings. In press

Antihistamines



- Enriched exposure insights

Powell G, et al. Enriching real-world evidence insights through QSAR based exposure definitions. *Value Health*. In press

What is deep learning?

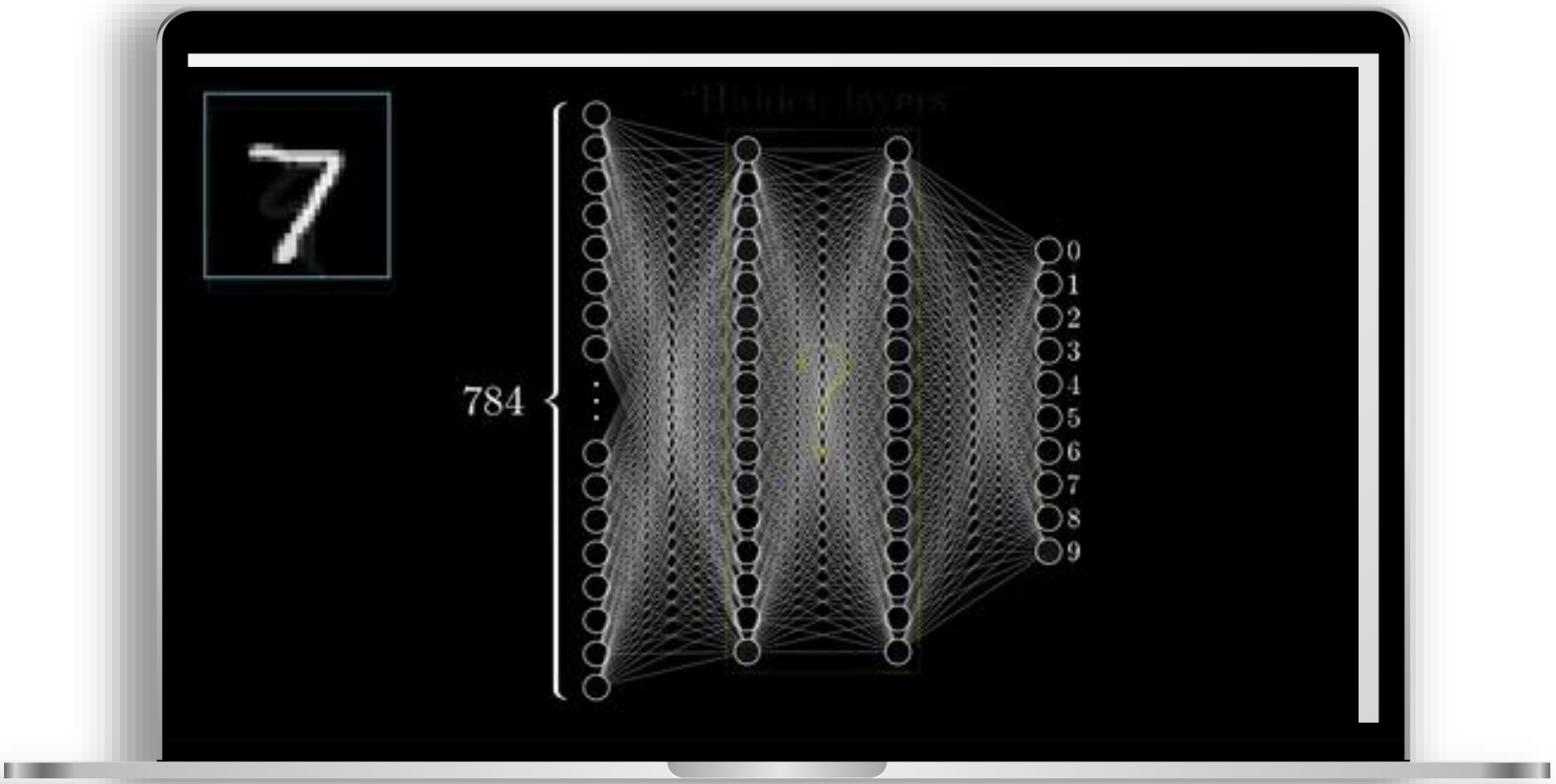


Image source: 3 Blue 1 Brown: https://www.youtube.com/watch?v=aircAruvnKk

Use of LLM in Pharmacovigilance –example 1

- FACTA+ used to extract signs and symptoms of listed AEFIs from MEDLINE for COVID-19 vaccines
- Tested ChatGPT Mapping of signs and symptoms retrieved from FACTA+ with PTs from MedDRA.
- Accuracy of GPT-3.5 was 78% for correct assignment of MedDRA PTs from signs and symptoms (kappa 1 across the 10 tests

Dong et al Optimizing Signal Management in a Vaccine Adverse Event Reporting System: A Proof-of-Concept with COVID-19 Vaccines Using Signs, Symptoms, and Natural Language Processing. Drug Safety. In Press

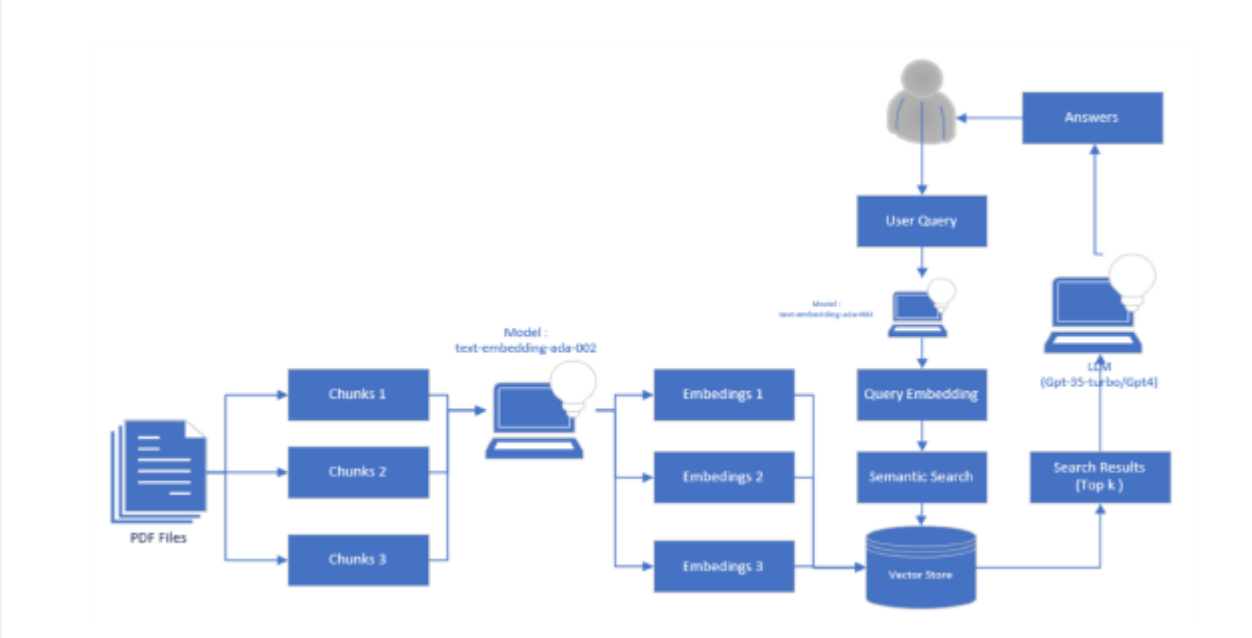
GSK

Use of LLM for Safety – example 2

Can LLM usefully and accurately provide information from a user guide?

Prompt driven interrogation of Sand-boxed environment

of Chat-GPT 4 with global user manuals as well as some country specific reference guides (596 pages)



- Painter JL et al 2023. Enhancing Drug Safety Documentation Search Capabilities with Large Language Models: A User-Centric Approach.
- In 2023 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE.IEEE CPS proceedings. In Press

Use of LLM for Safety – example 2

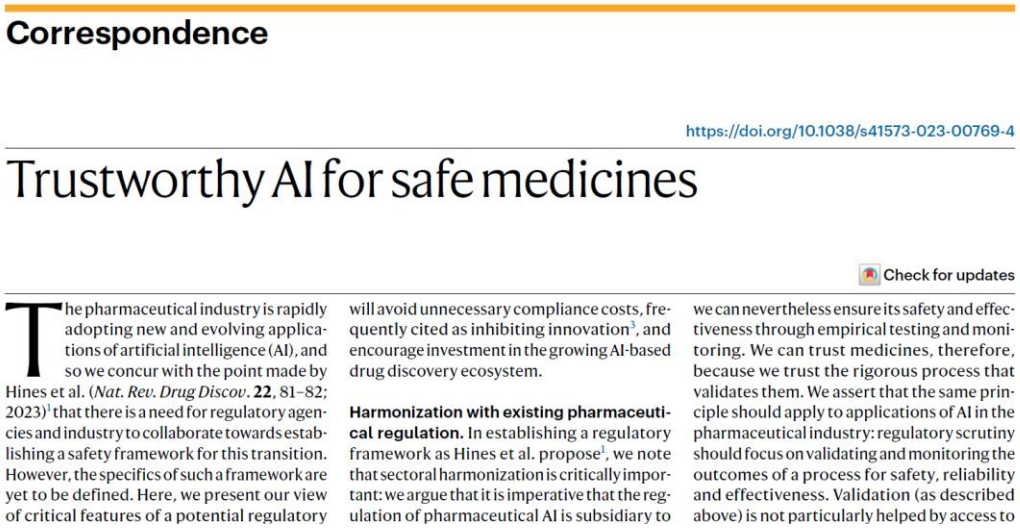
- Small study (22 core questions, 56 total variants), nevertheless results promising
- Test set of questions designed to: 1. Confirm Understanding, 2. Look for Guidance and Advice 3. Describe and Summarize and 4. Respond appropriate to nonsensical or Out of Context inputs
- Good consistency of answers: when prompted twice with the same question 73% of the LLM's responses were consistent
- Describing and summarizing procedures: LLM excelled, garnering the highest scores.
- re guidance and advice questions the LLM consistently recommended that users seek more specific guidance from their managers when it was uncertain about response.
- LLM demonstrated exceptional performance when confronted with nonsensical questions
  - No hallucination seen in study
- Low-scoring LLM answers were with clearly hard-to-retrieve user manual information
  - e.g. data within unreadable tables or figures, or complex situations with information spanning documents or specific guidance needs: output often lacking context or failing to cover varying scenarios, such as those between vaccines and drugs or AE source differences (e.g., clinical trials vs. spontaneous reports)
  - When guidance related to specific chapters, accuracy suffered due to mixed titles and chapter numbers.
- Qualitative analysis favoured concise and simple questions with single-document references. Nonsensical prompts were well-handled, with no attempt to generate a response.

Painter JL et al 2023. Enhancing Drug Safety Documentation Search Capabilities with Large Language Models: A User-Centric Approach.

In 2023 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE.IEEE CPS proceedings. In Press

GSK

Routine usage of AI in Pharmacovigilance



Correspondence

https://doi.org/10.1038/s41573-023-00769-4

Trustworthy AI for safe medicines

Check for updates

The pharmaceutical industry is rapidly adopting new and evolving applications of artificial intelligence (AI), and so we concur with the point made by Hines et al. (*Nat. Rev. Drug Discov.* **22**, 81–82; 2023)[1] that there is a need for regulatory agencies and industry to collaborate towards establishing a safety framework for this transition. However, the specifics of such a framework are yet to be defined. Here, we present our view of critical features of a potential regulatory will avoid unnecessary compliance costs, frequently cited as inhibiting innovation[3], and encourage investment in the growing AI-based drug discovery ecosystem.

Harmonization with existing pharmaceutical regulation. In establishing a regulatory framework as Hines et al. propose[1], we note that sectoral harmonization is critically important: we argue that it is imperative that the regulation of pharmaceutical AI is subsidiary to we can nevertheless ensure its safety and effectiveness through empirical testing and monitoring. We can trust medicines, therefore, because we trust the rigorous process that validates them. We assert that the same principle should apply to applications of AI in the pharmaceutical industry: regulatory scrutiny should focus on validating and monitoring the outcomes of a process for safety, reliability and effectiveness. Validation (as described above) is not particularly helped by access to

- Need a risk-based regulatory framework that implements proportionate precautionary measures to enable responsible innovation
- AI applications should not increase overall risk relative to relevant human benchmarks.
- Industry-wide risk-based framework should not require regulatory access to the underlying algorithms and datasets, particularly considering more effective alternatives.
- External stakeholders might still need some insight into datasets to ensure that they are methodologically sound, but this is possible through transparency tools such as datasheets and summaries
- Regulatory scrutiny for AI in PV should focus on validating and monitoring the outcomes of a process for safety, reliability and effectiveness.
- Sectoral harmonization is critically important
- Regulatory mechanisms must also be suitably agile to keep pace with the evolution of pharmaceutical AI.

Ref Stegmann et al

AI, artificial intelligence; ML, machine learning

Much interesting innovation for AI in PV –some examples

- Ball, R. and Dal Pan, G., 2022. “Artificial intelligence” for pharmacovigilance: ready for prime time?. Drug Safety, 45(5), pp.429-438.
- Lee, S., Kim, S., Lee, J., Kim, J.Y., Song, M.H. and Lee, S., 2023. Explainable Artificial Intelligence for Patient Safety: A Review of Application in Pharmacovigilance. IEEE Access.
- Pariente, A., Micallef, J., Lahouegue, A., Molimard, M., Auffret, M., Chouchana, L., Denis, B., Faillie, J.L., Grandvuillemin, A., Letinier, L. and Pierron, E., 2023. What place for intelligent automation and artificial intelligence to preserve and strengthen vigilance expertise in the face of increasing declarations?. Therapies, 78(1), pp.131-143.

GSK

Further reading on examples discussed here

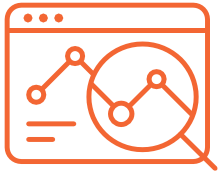
- Al-Azzawi F et al. 2023 Developing an artificial intelligence-guided signal detection in the Food and Drug Administration Adverse Event Reporting System (FAERS): A proof-of-concept study using galcanezumab and simulated data. Drug Safety. 46 (8), 743–751
- Bate A, Stegmann JU. 2023 Artificial intelligence and pharmacovigilance: what is happening, what could happen and what should happen? Health Policy and Technology. 12(2), 100743
- Bate A, Luo Y. Artificial Intelligence and Machine learning for safe medicines. Drug Saf. 2022;45:403-405.
- Dong G et al, Optimizing Signal Management in a Vaccine Adverse Event Reporting System: A Proof-of-Concept with COVID-19 Vaccines Using Signs, Symptoms, and Natural Language Processing. Drug Safety. In Press
- Kjoersvik O, Bate A. Black Swan Events and Intelligent Automation for Routine Safety Surveillance. Drug Saf. 2022;45:419-427.
- Kompa B, et al. Artificial Intelligence Based on Machine Learning in Pharmacovigilance: A Scoping Review. Drug Saf. 2022; 45:477-491.
- Painter JL et al. 2023. Enhancing Drug Safety Documentation Search Capabilities with Large Language Models: A User-Centric Approach. In 2023 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE.IEEE CPS proceedings. In Press
- Painter J et al. Frontiers In Drug Safety And Regulation. An Industry Perspective on the use of Machine Learning in Drug and Vaccine Safety. Frontiers in Drug Safety and Regulation. In Press
- Painter J et al. NLP and Machine Learning to Automate Identification of Suspected Medication Errors from Real World Unstructured Narratives. Value in Health. 26(6 Supplement) pp S281
- Stegmann JU et al. 2023. Trustworthy AI for safe medicines. Nature Reviews Drug Discovery, 22(10), pp.855-856.

GSK

Conclusions

- Machine learning has been used in assessing the safety of medicines ie Pharmacovigilance (PV) for many years
  - Routine capability of ML and advanced analytics currently shows value
  - More extensive ML inevitable for future next generation intelligent automation given the complexity of medicine/healthcare: trusted use is critical

- While the field is making inroads ML in PV is still immature and not widely used to maximum value
  - ML applied to specific problems/tasks within the PV lifecycle
  - Wider usage of other data streams for enrichment, contextualization and sometimes deeper insights is critical
  - Challenges are multiple and include the reliance/use of sparse data, validation frameworks and tools methods and processes as well as explicit and harmonized regulatory guidance



- To advance PV and use automation capability for patient safety, PV needs to be reconsidered fundamentally, not just superimpose AI/ML on antiquated systems, processes and frameworks – where only limited value and impact will be gained

ML, machine learning; PV, pharmacovigilance