AI in pharmacovigilance

A commercial buzzword or ready-to-use upgrade of the industry?

Martti Ahtola, Tepsivo 2022

List of Contents

Introduction

Artificial intelligence (AI) is a term that covers a wide variety of different technologies and concepts. Something that a layperson may see as AI is just a conditionality for a software developer. In this blog post, we go deeper in describing some of the technologies, how they can be utilized in pharmacovigilance, and what currently stands in the way.

We also try to quell the hype around AI in PV, as we know that many companies in the industry like to claim they utilize AI, while they often don't even understand what it means. It's important not to misinterpret the terminology and embrace the current limitations of AI and its, still, rare use within pharmacovigilance.

We are strong believers in AI replacing mundane tasks in PV, but we think it still has ways to go, and if we want to make real progress, we should be very clear about its current possibilities and not let premature hype shadow our judgment.

Application of AI should also be directed at the most critical areas where human effort has been overused the most, and tasks that are actually necessary to perform. Last thing we want to do is spend resources building sophisticated technology for tasks that make no sense to do in the first place.

The purpose of this article is to answer two main questions:

- What kind of AI technology is currently available?
- What are the potential implementations of AI in pharmacovigilance?

The low-hanging fruit for AI in pharmacovigilance

AI and automation have been hyped to take over pharmacovigilance for years. The main reason behind this hype is that currently most of the time in pharmacovigilance is used with repetitive administrative tasks that follow well defined rules, use defined formats and structures and even most PV experts are eagerly waiting to hand over these tasks to the robots and concentrate on the scientific assessment of the benefit-risk profile of the medicinal products.

Pharmacovigilance, and more specifically ICSR case processing has been also recognized by innovators and investors as an area that could be disrupted with the suitable technologies. Thus, there are already several companies that claim to provide artificial intelligence tools for pharmacovigilance and several others (such as the big consultancy companies) that aim to utilize in pharmacovigilance their AI and automation technologies which have been already implemented in finance and other business areas.

But, as most people working in the pharmacovigilance world know, very little artificial intelligence has actually been implemented in practice and while some new companies automate PV with smart tools (Tepsivo), AI is still very much at the hype stage.

Still, let's have a look at how AI already can be seen in some parts of pharmacovigilance, and more importantly, how we can utilize it in the future.

Different definitions of "AI" and areas of pharmacovigilance

There are different definitions and categorization for AI. AI covers various technologies / tools, and for every individual, the term can mean different things. Artificial intelligence includes, for example, statistical models, diverse algorithms, and self-modifying systems. Some would make a clear difference between AI and "automation" while at the same time others would include AI as a part of intelligent automation and hyperautomation. The main thing to understand for now is that AI consists of different methodologies and that those methodologies can be implemented in different ways.

These different methodologies are increasingly being implemented in all stages of a medicinal product's lifecycle, including clinical safety and post-authorization pharmacovigilance, but also for early research and development, simulations, marketing, sales, etc.

The main arguments for using AI within pharmacovigilance systems were initially based on faster processing of cases, elimination of human error, and standardization of processes.

Another obvious motivation is increasing the cost efficiency and gaining commercial edge compared to companies that rely on old-fashioned processes dependent on offshoring human labor; interestingly enough, old-school CROs and PV providers have little incentive to innovate as that would essentially dismantle their entire business model, rendering them obsolete (this will happen…).

Typical targets for AI in PV are the following:

- ICSR processing
- Aggregate reports
- Risk management
- Signal management
- Quality Management System (QMS)

These also happen to be the main areas of pharmacovigilance.

The below descriptions of AI technologies and potential applicability to pharmacovigilance has been customized and expanded with our own ideas from a comprehensive summary originally described in an excellent article by David John Lewis and John Fraser McCallum.

Artificial intelligence and Cognitive computing

Artificial intelligence (AI) is an all-embracing term for the simulation of human intelligence processes by computer systems. AI encompasses a wide range of technologies including following rules, reasoning (using rules to reach approximate or definite conclusions), learning, and self–correction.

Cognitive computing is the simulation of human reasoning in a computer system and often synonymous with AI. The goal of cognitive computing is to create automated systems that can solve problems with little or no human assistance using machine learning techniques.

AI and cognitive computing can be universal solutions and that's why they have potential in all the key areas of pharmacovigilance: ICSR processing, aggregate reports, risk management, signal management and QMS.

The goal is that in the future AI will be used for analysis and decision making in seriousness assessment, medical confirmation, medical review, and clinical evaluation either autonomously or together with a human expert. Using AI in the analysis helps to remove or at least reduce the biases and assumptions usually brought in by a human reporter, PV specialist and safety physician.

Neural networks

Artificial neural networks (ANN) were inspired by biological systems. They are typically organized in layers made up of thousands of interconnected nodes (inspired by the neurons of the brain). Data is presented to the network via an input layer which communicates to one or more hidden layers where the actual processing is done. These hidden layers then link to an output layer where the answer is surfaced. Examples include convolutional neural networks most commonly applied to analyze visual imagery and recurrent neural networks which can be applied to handwriting recognition and speech recognition.

Neural networks combine into a group of different application types of the same general idea; and it is a large group. This means that if you think about neural networks in general you can consider them to be a universally useful technology that has the potential to be implemented in all areas of pharmacovigilance. However, at the moment, there is a limited number of examples where this would already be happening in practice.

Machine learning, deep learning, and supervised learning

Machine learning (ML) is an application of AI that provides computerized systems with the ability to automatically learn and improve from experience without being explicitly programmed. Some implementations of machine learning use neural networks. ML focuses on the development of computer programs that can access data and use it to learn for themselves and adapt over time. The computer programs apply historic understanding to predict accurate outcomes from current inputs.

Deep learning is distinct from machine learning largely by depth of the neural network or the number of layers of the neural network.

Learning can be supervised, semi-supervised, or unsupervised. In supervised learning there is a human–annotated answer file ("ground truth") that is used to teach the machine learning algorithm. Unsupervised learning and reinforcement learning methods, where there is no "ground truth", may have utility in signal management, because they would avoid introduction of bias in identifying potential signals.

Machine learning has the potential to be used in ICSR processing, aggregate reports, risk management, signal management and QMS. Different applications of machine learning (specifically, supervised learning) have already been introduced for ICSR case processing and case assessment to some extent. The algorithm accesses the data from human PV specialists on the assessment of different kind of adverse event reports ("ground truth").

Text analytics and Text mining

Text analytics is the examination of large collections of written resources to generate new evidence or insight. Text analytics consists of a set of linguistic, statistical, and machine learning techniques that model and structure the information content of textual sources for useful information, exploratory data analysis, research, or investigation. Text analytics is roughly synonymous with text mining, while text analytics is now more commonly used in business settings. The goal of text mining is to discover relevant information in unstructured text, transforming or structuring this into data that can be used for further analysis or processes.

In pharmacovigilance processes, text mining can be used for ingesting an email directly into specific database fields or collation of relevant information in a clinical study report into an aggregate report. In larger scope of the product life cycle, biomedical text mining can be considered to be its own subset of bioinformatics and it is used to organize large sets of text data and to support scientific discovery. There are initiatives that aim to provide semantic cues to machines to answer specific questions included in the text while still protecting the content's access and licensing rights.

Sentiment analysis and Advanced analytics

Sentiment analysis is the contextual identification and extraction of meaning from text. It utilizes deep learning to understand intentions and reactions, and determine if an expressed opinion is favorable, unfavorable, or neutral, and to what degree.

Advanced analytics is the automated or semi-automated analysis of data using sophisticated tools such as machine learning, neural networks, and data mining to discover deeper insights, make predictions, or generate recommendations beyond those of traditional business intelligence.

Predictive analytics is a specific branch of advanced analytics that utilizes current and historical data to draw inferences to forecast activity, behavior, and trends. It involves applying statistical analysis techniques, and machine learning to data sets to create predictive models of a particular event happening.

An example of sentiment analysis in pharmacovigilance could be the assignment of reporter causality assessment relating to an adverse event for an administered medicinal product.

Predictive analytics would be extremely useful for authorities performing signal detection and for marketing authorization holders analyzing the available data for continuous safety-benefit profile monitoring and writing aggregate reports.

Use of analytic technology may be able to reduce the burden on human resources to isolate safety-relevant information from large documents, such as clinical or pre-clinical study reports, including those from outside sources. This would allow safety personnel to focus on the impact to patient safety (signal management, aggregate reporting) and reduce the administrative burden of managing complex documents.

Speech recognition

Speech recognition is an area of computer science that has been developed since the 70s by companies such as IBM. The area has sprung many other technologies such as machine-based trading which has largely replaced the classic human decision-based trading on Wall Street. Speech recognition has now become part of everyday life with Siri, Alexa, and voice dictation of messages and subtitles in YouTube.

Optical character recognition (OCR) is a technology that recognizes characters within a digital image. It is commonly used to recognize text in scanned documents. While OCR was designed for printed text, it can be used to verify handwritten text, as well.

The next step from OCR is Natural language processing (NLP). NLP helps computers understand human language, aiding interactions with humans in their own language and scaling language-related tasks. NLP can extract text from unstructured sources, interpret it, determine sentiment, and understand importance to create meaning.

Machine translation (MT) is the application of computers to the task of translating texts from one natural language to another. There are several dozens of services and technically different options for machine translation available.

Speech-to-text uses machine learning technologies to enable the recognition of spoken language and conversion of this into text. Speech synthesis (text-to-speech) is the use of computer systems to produce artificial human speech which is understandable to humans in natural language. Text-to-speech can be used for chatbot technology which will be discussed later on.

OCR is a commonly suggested (and already used) technology in the "AI" solutions for adverse event reporting to read medical articles, emails, CIOMS, MedWatch and other study specific forms. As a natural continuation, NLP can be used to recognize the connection between the "read" words in these texts highlighting adverse events, product, and patient information. In case processing, machine translation would be used when a report needs to be translated from the native language to English or report going to a local authority needs to be translated from English to the local language.

Using different combinations and integrations of these technologies, they have the potential to simplify and standardize the intake of ICSR data into a

PV system. It is still unclear how speech synthesis or recognition could be successfully integrated but these technologies could be employed to gather safety data from patients or prescribers with real-time querying to improve completeness of initial data capture and reduce follow-up burden in patient support programs and/or medical information centers.

Machine vision and Image recognition

Machine vision refers to many different technologies, software, and hardware. Machine vision is the ability of a computer to mimic sight and recognize objects to enable decisions or additional processing. Examples of machine vision include OCR and the interpretation of diagnostic test results.

Image recognition is the use of cameras, machine vision, and AI to enable a computer system to identify objects, places, people, and writing in static and video images.

Combination of machine vision and training related machine learning technologies have been part of everyday internet usage with reCAPTCHA since 2007. Google's version of CAPTCHA is used to teach company's AI technologies, for example teaching their self-driving cars to recognize cars, traffic signs and bicycles. While the public and legislators still doubt the self-driving technologies, it has been scientifically proven that image recognition technologies are already better at medical image analysis than experienced specialist physicians.

While machine vision's applications in pharmacovigilance might be limited to reading hand-written reports, recognizing attachment files of ICSRs or pictures in medical literature for literature screening, it has an important role in early diagnosis of disease.

Natural language generation

Natural language generation (NLG) is a computer process that automatically transforms structured data into a written or unstructured narrative. For any NLG software to produce human-ready narrative, the format of the content must be outlined (through templates, rules-based workflows, and intent-driven approaches) and then fed with structured data from which the output is created.

Natural language generation already has example implementations in software that produce ICSR narratives, but this could be further expanded to generation of periodic safety reports such as PSURs and DSURs, or for signal management other regular data-driven reports.

Autonomous software, RPA, and Desktop automation

Autonomous software is a software entity that carries out operations on behalf of a user with a degree of independence, employing some knowledge or representation of the user's goals or desires.

Robotic process automation (RPA) utilizes software "bots" to perform traditionally manual activities comprising high-volume, repetitive, rule-based processes involving structured data, such as adverse event case processing. RPA mimics execution of the repetitive activities without intervention or assistance.

Desktop automation is automation within a computer desktop to provide assistance or guidance to a human resource upon demand. Desktop automation can perform activities such as copying and pasting information, data entry, and opening applications. These activities occur on an employee's desktop and can be initiated by one or a combination of steps, such as a button click or switching tabs.

Bots are programs which carry out RPA. Bots work 24/7, at machine speed, without pausing, and are fully compliant with the process. Changes can be implemented instantly without training. Bots are scalable to suit the process. A variation of RPA is a smartbot, which is enriched by AI.

Chatbots are bots which conduct a conversation via audio or text methods and designed to convincingly simulate human conversation. Some chatbots are simple in operation, while others use NLP. For example, Google is offering simple chatbot programming for everyone. In its simplest form chatbots can be "if-then" operations and programmed in matter of hours.

RPA can be used for building automated processes in situations where simple automation and integration is not feasible. For example, RPA can be used for scraping data from local journals and other data sources such as regulatory authorities that do not offer API, RSS or other convenient way for acquiring data.

Automation technologies, while varied, are easily integrated to operate standardized workflows, which are currently heavily human-oriented. Orchestrated design of these workflows, combined with other advanced technologies, have the ability to mitigate manual, error-prone, and repetitive administrative tasks in ICSR management.

Blockchain

A blockchain is a continuously growing list of records, called blocks, which are linked and secured using cryptography (study of techniques for secure communication). By design, a blockchain is resistant to modification of data and it records transactions between two parties efficiently and in a verifiable and permanent way.

Blockchain technology has been widely adopted in financial systems and is used for tracking, tracing, auditing, and monitoring transactions. Blockchain can be used to operate and monitor a co–licensing contractual agreement between two different legal entities in which all transactional data are managed within the blockchain architecture. The implication is that all aspects of the partnership affecting the PV system are contained within a single protected environment, traceable, and readily available for audit or inspection. Blockchain technology could be used for many purposes in pharmacovigilance and it is an ideal technology considering the requirements on 21 CFR part 11 compliance.

First come to mind the contractual relationships such as SDEA, licensing agreements, and reconciliation between the parties. If this idea is expanded further, blockchain could be even used for submission and migration of ICSRs, for maintaining and updating aggregate reports, to provide evidence for audits or maintaining product's information (replacing eCTD).

However, there could be some resistance from the QA departments. Blockchain is by default open-source technology and as such, very transparent. Most companies' QA would not approve any open source-based software to be used in the company. Also, while there is a huge hype around blockchain, it might be a challenge to find solution providers that are already approved vendors or that could be qualified as vendors. This is just our gut feeling, hopefully we are incorrect and there are those QA / PV teams that are willing to go the extra mile to validate these new technologies.

Future of pharmacovigilance

Case processing is at the core of today's pharmacovigilance, and it is a rule-based process that ends with records in a safety database with well-defined fields with a lot of potential for automation. One thing to take into consideration is that while much of the automation efforts and AI implementation is focused on ICSRs, the relative contribution of ICSRs may be diminishing in the next 10 years, as other more robust sources of real-world data (RWD) and emerging methods become available.

The goal of collection of ICSRs is having as much information about the safety-benefit profile of the medicinal product as possible. However, ICSRs are in fact not a very good data source.

Firstly, only a fraction of adverse events experience by the patients gets reported to the authorities or the marketing authorization holders. For example, right now, in 2022, the regulatory authorities are still burdened by the sudden doubling in the number of reported ICSRs in 2021 due to COVID-19 vaccines and the huge number of reports from patients caused by the media attention.

However, for instance, the Finnish national competent authority Fimea received an adverse event report for about 0,15% administered COVID-19 vaccine doses. Think about yourself and the people close to you. Do you know anyone who had a sore arm and/or flu like symptoms after the vaccination?

COVID-19 vaccines were an extreme example due to the media attention and the adverse event reporting was probably much more frequent than with boring everyday medicine. That should give us an idea what percentage of adverse events is reported. Hype behind RWD and real-world evidence (RWE) is that all of us would have the possibility to report our health data in much easier ways, including scenarios where our wearable devices (or implants) submits all our health data to the authorities automatically.

Secondly, while to someone performing data entry it might seem that ICH E2B R3 (and even R2) have an endless number of fields and possibilities to include data, ICSR still are quite outdated and limited data source by modern standards. The data collected by your watch and phone even today probably could give a much better picture of your health, even though you would not be purposefully using them for health monitoring.

This is the reason all big tech companies are investing heavily in the healthcare sector. Digital health products and medical devices gathering real-world health data will probably be the not-so-distant future of pharmacovigilance. When the health data is collected in this manner, in large quantities, and without human interference, machine learning and other related AI techniques can become truly useful.

Conclusion

As seen in this blog post, artificial intelligence can refer to a large set of different technologies and different kind of combinations of those technologies. Out of those technologies, machine learning and its different subcategories is probably closest to the “thinking machine” idea that most of think of when AI is discussed in the pharmacovigilance data processing setting. However, to be fully useful, machine learning should be combined with other techniques such as text production.

At the moment, most of the artificial intelligence technologies for pharmacovigilance are still in the “could be used” or “will be used” state, but we agree with the hype, there is a huge amount of potential.

The biggest challenges are probably the strict regulation (for example, FDA wants to classify AI and ML as medical devices) and the lack of technological know-how in the healthcare industry. These two are actually the same thing: the health authorities lack the technological knowledge, as we have pointed out in several of our blogs, which leads to unfitting regulation. This then leads to the situation that the healthcare industry is using 25-year-old technology which is already borderline dangerous for patient health, in our opinion.

After these bitter statements, it should be highlighted that the overall feeling should be of excitement. These technologies can and will change our health and other parts of our lives in the near future and at least we at Tepsivo are happy to be part of this transformation.

Any questions?

If you need local Pharmacovigilance expertise, we are here to support you, anywhere in the world.

Visit our website www.tepsivo.com to learn more about our unique approach to PV services, or contact us directly at info@tepsivo.com.

Thank you for reading!
Martti Ahtola, 2022

tepsivo