



Industry Perspective on Artificial Intelligence/Machine Learning in Pharmacovigilance

Raymond Kassekert¹ · Neal Grabowski² · Denny Lorenz³ · Claudia Schaffer⁴ · Dieter Kempf⁵ · Promit Roy^{6,7} · Oeystein Kjoersvik⁸ · Griselda Saldana⁹ · Sarah ElShal¹⁰

Accepted: 17 February 2022
© TransCelerate BioPharma Inc. 2022

Abstract

TransCelerate reports on the results of 2019, 2020, and 2021 member company (MC) surveys on the use of intelligent automation in pharmacovigilance processes. MCs increased the number and extent of implementation of intelligent automation solutions throughout Individual Case Safety Report (ICSR) processing, especially with rule-based automations such as robotic process automation, lookups, and workflows, moving from planning to piloting to implementation over the 3 survey years. Companies remain highly interested in other technologies such as machine learning (ML) and artificial intelligence, which can deliver a human-like interpretation of data and decision making rather than just automating tasks. Intelligent automation solutions are usually used in combination with more than one technology being used simultaneously for the same ICSR process step. Challenges to implementing intelligent automation solutions include finding/having appropriate training data for ML models and the need for harmonized regulatory guidance.

✉ Neal Grabowski
neal.grabowski@abbvie.com

¹ GlaxoSmithKline, Global Safety, Upper Providence, PA, USA

² AbbVie, Pharmacovigilance and Patient Safety Business Process Office, North Chicago, IL, USA

³ Bayer AG, Medical Affairs and Pharmacovigilance, Pharmaceuticals, Berlin, Germany

⁴ Merck Healthcare, Case and Vendor Management-Global Patient Safety, Darmstadt, Germany

⁵ Genentech, A Member of the Roche Group, South San Francisco, CA, USA

⁶ Novartis, Chief Medical Office and Patient Safety, Novartis Global Drug Development, Dublin, Ireland

⁷ Trinity College, Dublin, Ireland

⁸ MSD, R&D IT, Prague, Czech Republic

⁹ Amgen, Pharmacovigilance Operations, Los Angeles, CA, USA

¹⁰ UCB, IT Patient Safety, Brussels, Belgium

Key Points

Pharmacovigilance organizations are very interested in and moving rapidly with planning, piloting, and production implementation of intelligent automation solutions that automate tasks (rule-based automation) and/or mimic human-like data interpretation and decision making (machine learning [ML] and artificial intelligence).

Multiple technologies can be applied simultaneously to the same process step.

Implementation of intelligent automation solutions faces challenges regarding quality training data for ML models and regulatory guidance.

1 Introduction

Intelligent automation, as defined by Lewis and McCaullum, including machine learning (ML) and artificial intelligence (AI), has started changing the way safety and pharmacovigilance (PV) professionals work to process and analyze data in support of decision making [1].

Intelligent automations are gaining traction as sponsors aim to improve efficiency, quality, timeliness of Individual Case Safety Report (ICSR) processing, and consistency of data for analysis and decision making, with the ultimate goal to benefit patients through more rapid identification and communication of product safety information.

In 2018, TransCelerate's Intelligent Automation Opportunities (IAO) in Pharmacovigilance Initiative (IAO team) coalesced around the vision of enabling organizations to manage and process the growing volumes of ICSRs efficiently through application of sophisticated technology [2]. The TransCelerate IAO team began evaluating risks and barriers to PV adoption and implementation of intelligent automation with a vision of identifying technology that could enable organizations to realize the potential benefits, such as improving scalability, improving efficiency, and quality [2]. Since its inception, the IAO team has published on potential use cases [1], the current state of automation within ICSR processing as of 2019 [3], and validation considerations [4] to enable value realization. IAO chose ICSR processing as the first area of study due to the disproportionate number of resources, human and financial, expended by PV departments compared with the interpretation and analysis of the data generated during the intake, processing, and submission steps.

This paper presents the current industry perspective on the role of intelligent automation in pharmacovigilance supported by recent TransCelerate survey data (conducted in 2021) and recently published literature. Furthermore, we discuss the challenges for developing and implementing advanced technologies, including (1) availability of representative training data and (2) the current regulatory environment.

Through the 2021 survey, several key themes emerged regarding the current state of intelligent automation in PV.

- Rule-based automations have long been used in ICSR processing, and PV organizations continue to embrace newer rule-based technologies such as robotic process automation (RPA) alongside other structured programming techniques in the ICSR process.
- There is continued interest in the implementation of emerging cognitive technologies such as ML, with increased activity compared with 2019. This activity is often in combination with other technologies or human effort.
- Implementers are stacking multiple intelligent automation technologies in addition to single-point solutions to realize the benefits beyond what traditional, rule-based technologies have allowed.

In addition to these key themes, there are other areas for consideration in the implementation of intelligent automation.

- Implementors of intelligent automation need to plan for potential challenges in curating suitable training data.
- Risk perceptions are changing, with respondents citing lower implementation risks. This change may be due to experience gained on prior pilot studies and implementations and may be despite challenges in generating voluminous, representative, and labeled training data and an emerging regulatory environment. Sponsors are proactively engaging with regulators on expectations when implementing IA.

This paper posits that intelligent automation will continue to transform ways of working within PV and more broadly within research and development (R&D) organizations faced with increasing volumes and velocity of data to expedite drug development and marketing approval.

2 Transcelerate Survey Methodology

The IAO team surveyed TransCelerate's Member Companies (MCs) in 2019, 2020, and 2021, with 15/19 (79%) MCs participating in 2019, 20/21 (95%) MCs participating in 2020, and 18/20 (90%) MCs participating in 2021. However, the 2019 survey measured attitudes about the ICSR process (i.e., the current manual effort of each process step and the perceived risk and expected benefit from automating each step) [3], and the 2020 survey collected benchmark technology implementation data based on the ICSR process (i.e., the type and maturity of automations implemented within each ICSR process step) [5]. The 2021 survey, discussed in this paper, sought to replicate and extend these results by combining elements of the 2019 and 2020 surveys into one follow-up survey. To facilitate consistency between the surveys, the same example ICSR process definition was used as described by Ghosh et al. [3], and the analysis is based on the three main processing blocks of intake, processing, and reporting.

In each case, the IAO team developed and distributed the survey to identified contact points within each TransCelerate MC. The contact points collated responses from their respective organizations and returned the completed survey forms to a single point of contact (POC) for IAO, a third-party consultant who serves as project manager for the IAO team. The data collection POC aggregated the data for analysis by the IAO technology subteam. The data were blinded and aggregated so that responses from individual MCs could not be discerned. To account for changes in the number of responses from one year to the other, the survey

response data were further processed to analyze the relative use of technology for the ICSR processes so trends could be identified. Survey responses from MCs were only shared with the TransCelerate POC, who ensured anonymization of each MCs name and aggregated the data without any company-identifying information remaining. A merger between two MCs resulted in fewer potential respondents in 2021 compared with 2020.

To aid in the analysis of the data, several visualizations were produced from the aggregated survey responses. Colors of varying intensity were used to show movement in responses between the survey. The stronger the movement, either increasing or decreasing, the deeper the color saturation, resulting in darker cells within the visualizations.

Survey forms from 2019, 2020, and 2021, including definitions of the ICSR process steps, intelligent automation technologies, and assessment categories, are provided in the electronic supplementary materials (ESM). Definitions used within the IAO team for the technologies

discussed in the surveys and this analysis are also included in the ESM. Several themes were identified through the analysis of survey data.

The TransCelerate MC survey results were further supplemented with a review of recently published literature conducted in PubMed from January 2019 to August 2021. Literature was retrieved using the search terms ‘pharmacovigilance automation’ and ‘pharmacovigilance machine learning’ and all abstracts were reviewed to select articles for full-text review.

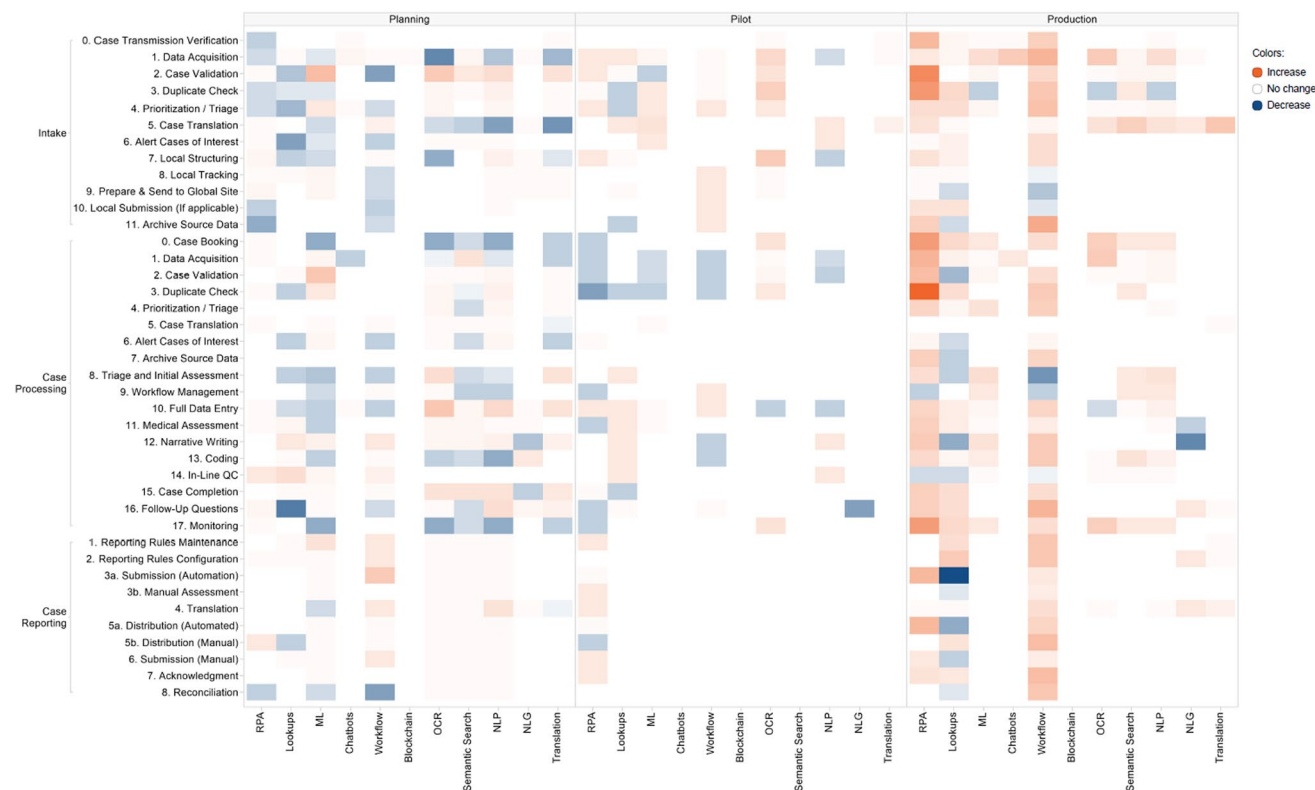


Fig. 1 Changes in TransCelerate’s Technology Survey from 2020 to 2021. In 2020 and 2021, TransCelerate member companies were asked to assess their activities in the adoption of intelligent automation technologies for the ICSR management process based on the planning, piloting, and implementation phases. ICSR management processes are represented on the y-axis and automation technologies used in the survey are represented on the x-axis. Results on the x-axis are grouped by the planning, pilot, and production phases. Blue indicates fewer responses when 2021 results were compared with 2020

results, and orange indicates more responses in the 2021 survey compared with 2020. Deeper/more intense colors indicate more difference in responses in 2021 compared with 2020. Response values ranged from - 14% (darkest blue) to + 35% (darkest orange), with white indicating 0% change. ICSR Individual Case Safety Report, QC quality check, RPA robotic process automation, ML machine learning, OCR optical character recognition, NLP natural language processing, NLG natural language generation

3 Identified Themes

3.1 Theme 1: Continued Development and Growth of Rule-Based and Other Structured Programming

In comparing the 2020 technology survey results with the 2021 responses (Fig. 1), we observed an implementation trend in the movement from planning or piloting into production. This trend is particularly observable for rule-based automations such as RPA, lookups, and workflows. In Fig. 1, a decrease in implementation is displayed in blue, whereas red-colored items indicate an increase. The color saturation/intensity is providing a measure representing the amount of change. For example, the duplicate check process step within the case processing block was largely piloted in 2020 with rule-based technologies such as RPA, lookups, and workflow; however, in 2021 it was moved into production for several TransCelerate MCs.

This trend is observed across the entire ICSR process (i.e., intake, processing, and reporting), although predominantly within intake and processing and less so in reporting. The increased adoption of rule-based automation within the intake block is possibly supported by existing, well-defined requirements for the tasks associated with

these process steps. For example, with an RPA bot, a process with fixed parameters can be automated (e.g., extraction of structured information from a defined form). Thus, the effort (i.e., time) spent on ‘local structuring’ can be reduced. As shown in Fig. 1, other process steps, such as duplicate check, appear to benefit from rule-based automations such as RPA.

Further industry survey data from PVNet[®] member organizations suggest broad interest in the adoption and implementation of rule-based automations such as workflow automation, labeling, and coding compared with the adoption of ML [6]. The PVNet[®] data predates TransCelerate’s IAO initiative and provides a contrast to show how perceptions of intelligent automation such as ML have evolved.

Lookups are another type of rule-based automation that is broadly used across all process steps. A lookup array is used in computer programming and holds values that would otherwise need to be manually decided (e.g., a list of adverse event terms that must always be considered ‘serious’). Since this technology is pervasive across the ICSR process, organizations may have already realized the potential value of rule-based automations. These diminishing returns may necessitate further innovation in the development and implementation of ML/AI to derive further benefit.

Figure 2 explores not only what process blocks are automated but also depicts the impact of intelligent automation

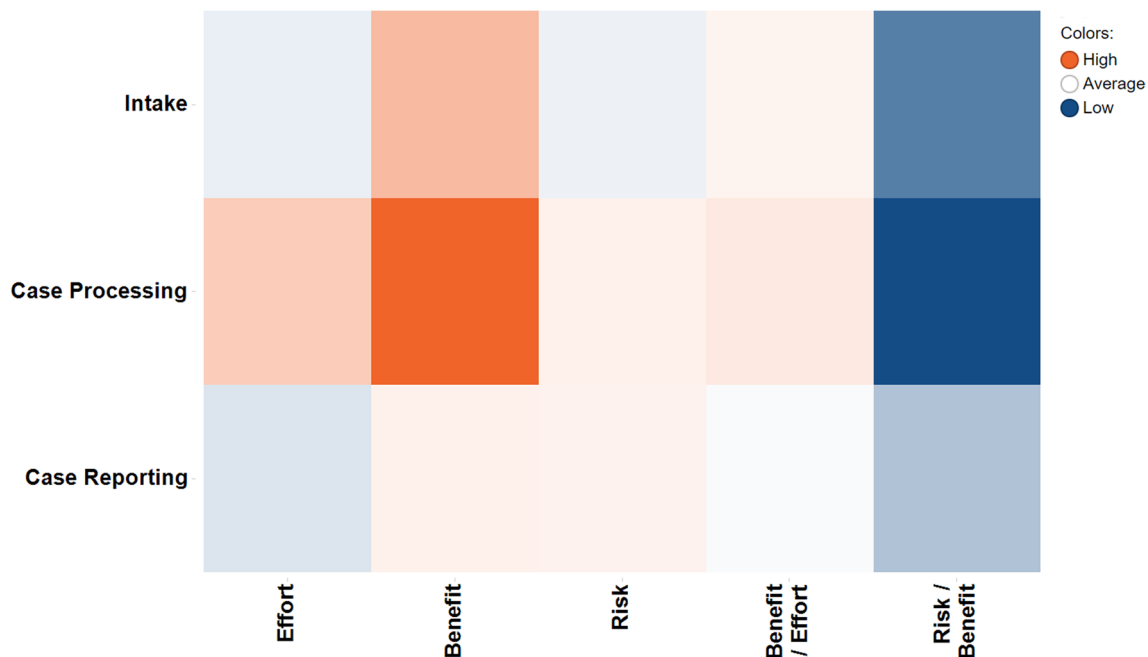


Fig. 2 Heatmap of effort-benefit-risk associated with the use of intelligent automation in ICSR processes. In the 2021 survey, the member companies were asked to assess, for each activity in the ICSR management, what risk, effort, and benefit is associated with the implementation of IA. This matrix displays the averages of process step data displayed in Table 1 of the associated effort, benefit, and risk

aggregated for the three main process blocks (intake, case processing and case reporting). The selected color schema indicates high values in red and low values in blue. The higher or lower the response was, the more intense the color is shown in the matrix. *ICSR* Individual Case Safety Report, *IA* intelligent automation



Fig. 3 TransCelerate’s Technology Survey Results 2021. In 2021, TransCelerate member companies were asked to assess their activities in the adoption of intelligent automation technologies for the ICSR management process based on the planning, piloting, and implementation phases. ICSR management processes are represented on the y-axis and automation technologies used in the survey are represented on the x-axis. Results on the x-axis are grouped by the plan-

ning, pilot, and production phases. The more a technology is used, the more intense the color is displayed in the matrix. Response values range from 0% (white) to 61% (darkest orange). *QC* quality check, *RPA* robotic process automation, *ML* machine learning, *OCR* optical character recognition, *NLP* natural language processing, *NLG* natural language generation

across the three ICSR process blocks of intake, processing, and reporting in terms of (1) the related effort associated with the task; (2) the risk associated with automation of that task; and (3) the benefit obtained through automation of that step. The same color scheme as in Fig. 3 was applied, with shades of red for the indication of high and shades of blue for a low association. Figure 2 displays the high perceived benefit associated with intake and case processing blocks, as indicated with a more intense red, compared with the medium perceived benefit for case reporting. This finding could be related to the high effort currently spent in intake and case processing without automation, and hypothetically could be saved using automation. The associated risk related to the automation of case intake is assessed as slightly lower than the associated risk for case processing, which could be explained by the fact that rules can be applied more efficiently during intake. In contrast, during case processing, the medical assessment-related activities are conducted under conditions requiring more complexity or presenting limitations to automation based on current, rules-based approaches.

A more granular look into the detailed ICSR process steps is shown in Table 1. This heatmap is an evolution from the previous publication, ‘Automation Opportunities in Pharmacovigilance: An Industry Survey’ by Ghosh et al. [3], and compares effort, benefit, and risk for the detailed process steps. As outlined in the caption for Table 1, this heatmap contains benefit and perceived implementation risk data from the 2019 survey; effort, benefit, and risk from the 2021 survey; and technology implementation data from the 2020 and 2021 surveys. While Fig. 1 focuses on the type of intelligent automation applied in the individual steps of the process, Table 1 visualizes effort, benefit, and risk per step and the percentage of MCs using automation at each step. For the ‘duplicate check’ step in the case processing block, the percentage of companies using intelligent automation increased from 25 to 78%. The significant progress in implementing automation, like RPA, could be attributed to the implicit logic and structure across the highly transactional ICSR process. Moreover, an RPA bot can be developed that mimics human interaction in safety database systems.

Table 1 Perceived levels of current effort and benefit versus risk from automation, automation opportunity, and level of applied automation per ICSR process step.

PROCESS STEP	2021 Current Level of Effort	2019 Level of Benefit from Automation	2021 Level of Benefit from Automation	2019 Level of Risk from Automation	2021 Level of Risk from Automation	2021 Automation Opportunity Score	% of Companies already applying Automation	2020 % of Companies already applying Automation	2021 % of Companies already applying Automation	Delta % of Companies already applying Automation
INTAKE										
0. Case Transmission Verification	MEDIUM	not collected	HIGH	not collected	MEDIUM	3.0	35%	56%	21%	21%
1. Data Acquisition	MEDIUM	HIGH	HIGH	MEDIUM	MEDIUM	3.0	40%	61%	21%	21%
2. Case Validation	MEDIUM	HIGH	MEDIUM	MEDIUM	MEDIUM	2.0	20%	61%	41%	41%
3. Duplicate Check	MEDIUM	HIGH	HIGH	MEDIUM	MEDIUM	3.0	30%	61%	31%	31%
4. Prioritization/Triage	MEDIUM	HIGH	MEDIUM	HIGH	MEDIUM	2.0	30%	50%	20%	20%
5. Case Translation	HIGH	HIGH	HIGH	HIGH	MEDIUM	4.5	25%	50%	25%	25%
6. Alert Case of Interest	MEDIUM	HIGH	MEDIUM	MEDIUM	MEDIUM	2.0	25%	33%	8%	8%
7. Local Structuring	MEDIUM	HIGH	MEDIUM	HIGH	MEDIUM	2.0	20%	33%	13%	13%
8. Local Tracking	MEDIUM	not collected	MEDIUM	not collected	MEDIUM	2.0	20%	33%	13%	13%
9. Prepare and Send to Global Site	MEDIUM	HIGH	MEDIUM	MEDIUM	MEDIUM	2.0	55%	50%	-5%	-5%
10. Local Submission (if applicable)	MEDIUM	HIGH	MEDIUM	MEDIUM	MEDIUM	2.0	20%	28%	8%	8%
11. Archive Source Data	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	2.0	45%	72%	27%	27%
PROCESSING										
0. Case Booking	MEDIUM	HIGH	MEDIUM	MEDIUM	MEDIUM	2.0	55%	83%	28%	28%
1. Data Acquisition	MEDIUM	HIGH	MEDIUM	MEDIUM	MEDIUM	2.0	45%	67%	22%	22%
2. Case Validation	MEDIUM	HIGH	HIGH	MEDIUM	MEDIUM	3.0	40%	61%	21%	21%
3. Duplicate Check	MEDIUM	HIGH	HIGH	MEDIUM	HIGH	2.0	25%	78%	53%	53%
4. Prioritization / Triage	MEDIUM	HIGH	HIGH	HIGH	MEDIUM	3.0	50%	78%	28%	28%
5. Case Translation	HIGH	HIGH	HIGH	HIGH	MEDIUM	4.5	5%	11%	6%	6%
6. Alert Cases of Interest	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	2.0	45%	50%	5%	5%
7. Archive Source Data	MEDIUM	HIGH	MEDIUM	MEDIUM	MEDIUM	2.0	50%	67%	17%	17%
8. Triage and Initial Assessment	MEDIUM	HIGH	MEDIUM	HIGH	MEDIUM	2.0	45%	50%	5%	5%
9. Workflow Management	MEDIUM	HIGH	MEDIUM	MEDIUM	MEDIUM	2.0	60%	67%	7%	7%
10. Full Data Entry	HIGH	HIGH	HIGH	HIGH	MEDIUM	4.5	45%	72%	27%	27%
11. Medical Assessment	MEDIUM	HIGH	HIGH	HIGH	MEDIUM	3.0	30%	39%	9%	9%
12. Narrative Writing	HIGH	HIGH	HIGH	MEDIUM	MEDIUM	4.5	65%	61%	-4%	-4%
13. Coding	MEDIUM	HIGH	HIGH	HIGH	MEDIUM	3.0	65%	72%	7%	7%
14. In-Line QC	MEDIUM	HIGH	MEDIUM	MEDIUM	MEDIUM	2.0	40%	39%	-1%	-1%
15. Case Completion	LOW	MEDIUM	MEDIUM	MEDIUM	LOW	2.0	40%	56%	16%	16%
16. Follow-Up Questions	HIGH	HIGH	HIGH	HIGH	MEDIUM	4.5	15%	39%	24%	24%
17. Monitoring	LOW	HIGH	MEDIUM	MEDIUM	MEDIUM	1.0	35%	44%	9%	9%
REPORTING										
1. Reporting Rules Maintenance	HIGH	HIGH	MEDIUM	HIGH	HIGH	2.0	15%	33%	18%	18%
2. Reporting Rules Configuration	HIGH	HIGH	MEDIUM	HIGH	HIGH	2.0	15%	33%	18%	18%
3a. Submission (Automation)	MEDIUM	HIGH	HIGH	MEDIUM	MEDIUM	3.0	80%	72%	-8%	-8%
3b. Manual Assessment	MEDIUM	MEDIUM	MEDIUM	MEDIUM	HIGH	1.3	25%	28%	3%	3%
4. Translation	HIGH	HIGH	HIGH	MEDIUM	MEDIUM	4.5	25%	33%	8%	8%
5a. Distribution (Automated)	LOW	HIGH	MEDIUM	MEDIUM	MEDIUM	1.0	65%	67%	2%	2%
5b. Distribution (Manual)	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	2.0	20%	44%	24%	24%
6. Submission (Manual)	MEDIUM	HIGH	MEDIUM	MEDIUM	MEDIUM	2.0	20%	39%	19%	19%
7. Acknowledgment	LOW	MEDIUM	MEDIUM	MEDIUM	MEDIUM	1.0	65%	72%	7%	7%
8. Reconciliation	MEDIUM	HIGH	MEDIUM	MEDIUM	MEDIUM	2.0	20%	39%	19%	19%

Heat map of perceived effort, benefit and risk of automation from the 2021 TransCelerate MC survey based on the average of the responses for each process step, represented at levels of low, medium and high. 2019 survey responses for benefit and risk, originally collected at different scales, were converted to low, medium and high values and are shown in comparison. Data on process steps not covered by the 2019 survey are marked as ‘not collected’. Automation opportunity scores are calculated as effort * benefit/risk (with low = 1, medium = 2, high = 3), with a higher score indicating a higher value opportunity. The table also shows the percentage of responding companies that were already applying automation for individual process steps in response to the 2020 and 2021 surveys, as well as the delta between the survey responses. *QC* quality check, *ICSR* Individual Case Safety Report, *MC* member company

3.2 Theme 2: Continued Interest in Emerging Cognitive Technologies Such as Machine Learning

According to the 2021 survey results, there is interest in emerging cognitive technologies such as ML, natural language processing (NLP), and Natural Language Generation (NLG) [Fig. 3], as indicated by frequent use in Planning. There is consistent interest in ML, NLP, and NLG planning from 2020 to 2021 (count of responses across all process steps for ML, NLP, and NLG, respectively, where survey response = ‘planning’: 100 to 100; 64 to 63; 16 to 15) [see also Fig. 1], even as pilot/production activity decreased (count of responses across all process steps for ML, NLP and NLG, respectively, where survey response = ‘pilot’ or ‘production’: 12/23

to 14/33; 7/24 to 6/30; 2/7 to 0/7). Furthermore, these gains were observed with fewer MC respondents in 2021 compared with 2020 (18 vs. 20). Significant planning for implementing these technologies was already observed in the previous survey results, which aligns with the rise of digitalization embraced by the pharmaceutical industry. Interest in automating these process steps remains constant, as automation decreases the effort for activities that do not require significant human perceptual or cognitive skills. As mentioned above, the results also show sustained growth in RPA. However, with increased growth in AI and ML implementation, it is now possible to automate tasks that are assumed to require human perceptual or cognitive skills, such as recognizing handwriting, speech, or faces; understanding language; planning; reasoning from partial or uncertain information; and learning. Technologies able

to perform tasks such as these traditionally assumed to require human intelligence are called cognitive technologies [1].

It was evident that companies are interested in using ML across the entire range of process steps, and that ML is in planning for almost all process steps, especially in Intake and Case Processing, as seen in the technology heat map (Fig. 3).

Focusing on the changes from the 2020 to 2021 surveys (Fig. 1), we can see progress with intelligent automation tools reaching production across many process steps, while many intelligent automation solutions remain in the planning or pilot phase.

3.3 Theme 3: Stacking of Technologies for Value

According to the 2021 survey results, there are no process steps in the example ICSR process flow, where companies had not taken any action to automate the process steps. Even for steps with no intelligent automation in production, respondents were planning or piloting some intelligent automation technology (Fig. 3). For all process steps, respondents have implemented some degree of intelligent automation. In these areas where intelligent automation was being implemented, respondents typically used multiple technologies in each step. In both the 2020 and 2021 survey, there were 38/40 process steps where some MCs reported implementing two or more intelligent automation technologies for the process step. Although a majority of MCs implemented a single technology for any given process step (70.1% in 2020 and 58.9% in 2021), some MCs have reported applying as many as six discrete intelligent automation technologies for a single process step. Comparing results from the 2020 and 2021 surveys indicates an increasing technology stacking trend, at least to a certain point, across all ICSR process steps among companies implementing at least two technologies for a given step: two technologies in production (20.1% in 2020 vs. 27.7% in 2021), three technologies (3.7% vs. 7.2%), four technologies (5.1% vs. 5.1%), five technologies (0.7% vs. 0.3%), and six technologies (0.3% vs. 0.8%) [see the full table in the ESM for more detail].

From the survey results, intelligent automation technologies are often not implemented in isolation, but instead, multiple technologies are increasingly being combined or stacked to produce a working solution.

Nearly every step in the example ICSR process has some workflow support. Workflow technologies are not necessarily new, and many commercial adverse event systems have extensive workflow capabilities. It would make sense intuitively that companies would combine other orchestration technologies in conjunction with workflow. This result is visible in our 2021 survey results (Fig. 3), showing RPA in production for 33/40 process steps. Workflow Orchestration

is also in widespread production. For every step where RPA is being applied, the respondents also report having workflow in production on those same steps. There are no steps in our survey where only one intelligent automation technology has been implemented to automate the process flow fully.

Optical character recognition (OCR) technology is another technology that appears in the automation plans for several process steps. In both the 2020 and 2021 survey, MCs reported having OCR in production for 11 process steps. Many companies receive source materials for ICSRs in portable document format (PDF). Where the PDF is computer-generated, extracting relevant text from the document through traditional programming methods is easily feasible. In many cases though, the PDF contains a rendered image of some structured format, such as a MedWatch or Council for International Organizations of Medical Sciences (CIOMS) form. In many cases, the PDF will be a rendering of a paper form that has been imaged. Both the 2020 and 2021 survey results show MCs using both OCR and NLP stacked or in combination in process steps where PDF images would be processed. For non-English text, automated translation can be used to obtain an English translation. NLP can classify the type of information in the document to facilitate downstream case-processing activities. MCs survey responses from both 2020 and 2021 show that for 11 process steps where NLP is in production, OCR is also in production. NLP can be used to identify demographic information such that the information relevant to the reporter can be identified separately from information relevant to the patient. The NLP classifier can be used to identify terms that should be coded using Medical Dictionary for Regulatory Activities (MedDRA), where adverse events can then be separated from medical history, indications for use, or concurrent conditions.

Within the ICSR process framework, several cases exist of ML and AI applied alongside orchestration technologies such as RPA or workflow automation. For example, ML algorithms trained on small sets of ICSRs from the US FDA Adverse Event Reporting System (FAERS) showed promise in classifying causality as certain, probable, possible, and unlikely [7]. Similarly, ML models showed promise in identifying adverse drug reactions (ADRs) from a patient-reported narrative text but fell short at assessing seriousness, a task that could be viewed as more complex [8]. While these technology evaluations are exploratory to date, further development and enhanced performance would undoubtedly benefit an RPA-orchestrated ICSR processing pipeline.

4 Discussion

There are several areas where further exploration is required, and implementors of intelligent automation should take note.

4.1 Training Data Challenges

As ML becomes a significant contributor to task automation within the pharmaceutical industry, additional emphasis needs to be put on data quality and data understanding. A supervised ML model requires representative training data with a valid ground truth established. In the pharmaceutical industry, this is often represented as historical datasets from existing systems that may come in various forms (e.g., sensor data, images, voice, text, and video). These data may not have been created to train future ML models and may lack the proper formatting or annotation. Within PV, ICSR processing may be an exception as companies have been extracting and interpreting millions of source documents into a structured database.

Representative training data are important as patterns from training data are generalized and then applied to new incoming data to make decisions based on similar problems or slight variations from data observed in the training dataset. If these patterns are inaccurate in the context of the new data, the decisions will be of low quality. This risk presents a challenge for companies if data were never collected for this purpose or never cleansed to accommodate such activities. Even within a single company's database, variations in ICSR data may exist regarding multiple languages, different conventions over time, or as the result of acquisitions and mergers. Other strategies to create a training dataset include voluminous manual curation by PV professionals, which can be costly and onerous. One novel alternative is crowdsourcing a training dataset that proved accurate, taking only a fraction of the time [9].

4.2 Perceived Risk in an Emerging Regulatory Environment

Survey data in Table 1 shows that from 2019 to 2021, the perceived level of risk from automation decreased in 36/38 (94.7%) evaluable ICSR process steps (two steps could not be compared as they were not collected in 2019). Duplicate check and manual assessment, defined as manual review of non-automatic submissions (e.g., automatable with rule-based technologies), showed higher risk from automation in the recent survey. This change in perceived risk may be attributable to the experience gained through previous pilots and production implementations.

Despite this gained experience, the emerging regulatory environment may present challenges to implementation progress such that many intelligent automation use cases in planning or piloting could experience unnecessarily long delays before advancing into production. Surprisingly, industry experts considered AI more mature in PV than in other R&D functions, such as regulatory and clinical operations, but still identified regulatory and compliance

concerns as one of the top reasons for not implementing AI [10]. As companies evaluate digital tools, they must consider and monitor the risk associated with AI use in certain areas of work.

In the absence of clear guidance on validation approaches, companies tend toward a conservative validation approach and carefully consider actions and approaches to ensure unforeseen events, model drift, and other risks are managed post-implementation. This approach has often accompanied proactive dialog with regulators and the consensus of acceptable quality levels, production monitoring, and risk management. This approach is further elaborated by Huysentruyt et al. in the proposed framework for validating intelligent automation systems, focusing on quality assurance and health authority engagement for acceptable performance measures [4].

The principles of the International Society for Pharmaceutical Engineering (ISPE) Good Automated Manufacturing Practices and the FDA Good Machine Learning Practices continue to be the foundation of standard-setting organization and health authority publications, including those of the Danish Medicines Agency (DKMA) [11], US National Institute of Standards and Technology (NIST) [12], and UK Medicines and Health products Regulatory Agency (MHRA) [13]. These documents signal a pragmatic approach toward AI validation focusing on training data, performance measures, and production oversight appropriate for using the system. We hope that AI validation expectations harmonize globally in an already complex operating environment burdened with numerous local requirements.

Other functions within the pharmaceutical industry that may not be subject to such rigid quality standards and inspections, such as drug discovery, have moved forward with AI implementation. As many as 88% of large-sized and 74% of mid-sized pharmaceutical companies used some form of AI somewhere within their organizations, compared with 50% of MCs in our survey who reported implementing AI within ICSR processing [10].

5 Conclusions

Automation has and will continue to change how PV organizations collect, process, analyze, and act on data. While rule-based automations are commonplace, they offer incremental efficiencies primarily by mimicking human transformation and processing of structured data. ML and other AI-based intelligent automations can genuinely be transformational and change how data enters and flows through PV organizations. Mockute et al. identified 51 decision points within ICSR processing that are candidates for AI [14], while countless others exist downstream in signal detection and management.

Intelligent automation technologies are not deployed in isolation and are complemented by rule-based automations to facilitate workflows and other ML models optimized for specific data points or tasks. While there is a vision for touchless case processing [15], the current state of ML will augment PV professionals by increasing their efficiency and effectiveness [7, 16].

Beyond ICSR processing, other PV processes, including AI literature review and social listening, were frequently reported, as in the pilot or planning stages [10]. Additional signal detection and evaluation uses include the ability to connect and synthesize evidence from multiple data sources across R&D from multiple data sources and synthesize evidence from molecule to patients [17]. The continued evolution of sponsor risk perceptions and regulator acceptance of validation approaches will allow greater future benefits from ML and other intelligent automation across PV domains.

Acknowledgements The authors worked as part of TransCelerate Biopharma Inc., which is funded by 20 organizations dedicated to making R&D, including pharmacovigilance monitoring and reporting, more efficient and effective in the pharmaceutical industry. The authors also gratefully acknowledge the support of the following TransCelerate working team members who contributed to the concepts described in this manuscript. We thank the late Sundeep Sethi of AbbVie for his early leadership of this working group and Bart Teeuw of UCB for his previous leadership; Kate Gofman of AstraZeneca and Andreas Duczmal of UCB for their participation in this working group, including their input on survey designs, previous hypothesis development, and earlier data analysis; and Clint Craun of TransCelerate for his continual project management support, survey distribution and response organization, and coordination with enabling TransCelerate functions.

Declarations

Funding This work was supported by TransCelerate BioPharma Inc.

Conflict of interest Raymond Kassekert, Neal Grabowski, Denny Lorenz, Claudia Schaffer, Dieter Kempf, Promit Roy, Oeystein Kjoersvik, Griselda Saldana and Sarah ElShal declare no conflicts of interest; however, all authors affiliated with pharmaceutical companies are employees and, in some cases, stockholders of the companies.

Availability of data and materials The surveys conducted in this study are provided in the ESM.

Code availability Not applicable.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Author contributions All authors contributed to the manuscript's conception and design, and read and approved the final manuscript. Survey distribution was done/coordinated by a third party, and survey data were collected and analyzed by a third party.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

- Lewis DJ, McCallum JF. Utilizing advanced technologies to augment pharmacovigilance systems: challenges and opportunities. *Ther Innov Regul Sci.* 2020;54:888–99. <https://doi.org/10.1007/s43441-019-00023-3>.
- TransCelerate. Intelligent automation opportunities in pharmacovigilance. <https://www.transceleratebiopharmainc.com/assets/intelligent-automation-opportunities-pharmacovigilance-solutions/>. Accessed 22 Jul 2021.
- Ghosh R, Kempf D, Pufko A, Martinez LFB, Davis CM, Sethi S. Automation opportunities in pharmacovigilance: an industry survey. *Pharm Med.* 2020;34:7–18. <https://doi.org/10.1007/s40290-019-00320-0>.
- Huysentruyt K, Kjoersvik O, Dobracki P, Savage E, Mishalov E, Cherry M, et al. Validating intelligent automation systems in pharmacovigilance: insights from good manufacturing practices. *Drug Saf.* 2021;44:261–72. <https://doi.org/10.1007/s40264-020-01030-2>.
- TransCelerate. Intelligent automation opportunities in pharmacovigilance. 2020 Technology Matrix. https://www.transceleratebiopharmainc.com/wp-content/uploads/2020/04/IAO_Technology-Matrix_April-2020.xlsx. Accessed 31 Jul 2021.
- Stergiopoulos S, Fehrle M, Caubel P, Tan L, Jebson L. Adverse drug reaction case safety practices in large biopharmaceutical organizations from 2007 to 2017: an industry survey. *Pharm Med.* 2019;33:499–510. <https://doi.org/10.1007/s40290-019-00307-x>.
- Kreimeyer K, Dang O, Spiker J, Muñoz MA, Rosner G, Ball R, Botsis T. Feature engineering and machine learning for causality assessment in pharmacovigilance: lessons learned from application to the FDA adverse event reporting system. *Comput Biol Med.* 2021;135:104517. <https://doi.org/10.1016/j.compbiomed.2021.104517>.
- Létinier L, Jouganous J, Benkebil M, Bel-Létoile A, Goehrs C, Singier A, et al. Artificial intelligence for unstructured healthcare data: application to coding of patient reporting of adverse drug reactions. *Clin Pharmacol Ther.* 2021;110:392–400. <https://doi.org/10.1002/cpt.2266>.
- Gartland A, Bate A, Painter JL, Casperson TA, Powell GE. Developing crowdsourced training data sets for pharmacovigilance intelligent automation. *Drug Saf.* 2021;44:373–82. <https://doi.org/10.1007/s40264-020-01028-w>.
- Lamberti MJ, Wilkinson M, Donzanti BA, Wohlhieter GE, Parikh S, Wilkins RG, Getz KA. Study on the application and use of artificial intelligence to support drug development. *Clin Ther.* 2019;41:1414–26. <https://doi.org/10.1016/j.clinthera.2019.05.018>.
- Danish Medicines Agency (DKMA). Suggested criteria for using AI/ML algorithms in GxP. <https://laegemiddelstyrelsen.dk/en/licensing/supervision-and-inspection/inspection-of-autho>

- rised-pharmaceutical-companies/using-aiml-algorithms-in-gxp/. Accessed 9 Aug 2021.
12. Schwartz R, Down L, Jonas A, Tabassi E. A Proposal for identifying and managing bias in artificial intelligence. Draft NIST Special Publication 1270. <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270-draft.pdf>. Accessed 1 Aug 2021.
 13. Nguyen-Van-Tam D. A.I. in Pharmacovigilance Case Processing—An MHRA Pharmacovigilance Inspector's Perspective. DIA Europe 2021. Virtual presentation, 16 March 2021.
 14. Mockute R, Desai S, Perera S, Assuncao B, Danysz K, Tetarenko N, et al. Artificial intelligence within pharmacovigilance: a means to identify cognitive services and the framework for their validation. *Pharm Med.* 2019;33:109–20. <https://doi.org/10.1007/s40290-019-00269-0>.
 15. Bate A, Hobbiger SF. Artificial intelligence, real-world automation and the safety of medicines. *Drug Saf.* 2021;44:125–32. <https://doi.org/10.1007/s40264-020-01001-7>.
 16. Danysz K, Cicirello S, Mingle E, Assuncao B, Tetarenko N, Mockute R, et al. Artificial intelligence and the future of the drug safety professional. *Drug Saf.* 2019;42:491–7. <https://doi.org/10.1007/s40264-018-0746-z>.
 17. Hauben M, Hartford CG. Artificial intelligence in pharmacovigilance: scoping points to consider. *Clin Ther.* 2021;43:372–9. <https://doi.org/10.1016/j.clinthera.2020.12.014>.