#### PERSPECTIVE



# The Role of Artificial Intelligence and Machine Learning in Polymer Characterization: Emerging Trends and Perspectives

Rick S. van den Hurk<sup>1,2</sup> · Bob W. J. Pirok<sup>1,2</sup> · Tijmen S. Bos<sup>1,2</sup>

Received: 17 January 2025 / Revised: 18 March 2025 / Accepted: 19 March 2025 / Published online: 4 April 2025 © The Author(s) 2025

#### Abstract

The application of artificial intelligence (AI) and machine learning (ML) is rapidly expanding and has begun to make a significant impact on polymer development and characterization. This perspective article explores the current state of AI in this field and highlights areas where its potential remains underutilized. While the optimization of polymer synthesis to achieve desired properties and the classification of polymer types are well-established, opportunities for AI integration in detailed characterization, analytical method development, and data processing remain largely untapped. Greater automation of the analytical laboratory, whether through dedicated algorithms or AI-driven solutions, will enable analytical chemists to focus more on addressing research questions and interpreting results, rather than on method development and routine measurements.

Keywords Polymer characterization · Machine learning · Artificial intelligence

# Introduction

In modern society, polymers play a pivotal role, with applications spanning everyday items (e.g., consumer electronics, photovoltaics, coatings, food packaging, etc.) to advanced materials (e.g., medical implants, space travel). To address present and future challenges, the development of sophisticated polymers with increasingly tailored properties is essential. While polymer synthesis offers significant flexibility in achieving diverse properties, chemists ultimately require reliable information to optimize the synthesis process. Generally, this information can be obtained through analytical methods, or physical property-based relations combined with trial-and-error experiments. Often, the physical properties of a polymer can be linked to its molecular structure, offering valuable insights into the structure–property relationship.

To streamline the research and development process and provide this information on the structure-property relationship, automated platforms have emerged as valuable tools for polymer synthesis [1-4]. The advent of machine learning (ML) has only contributed to this for cases where optimal polymer properties can be represented in a scoring framework. This score quantifies the desirability or the optimality of the polymer properties and can then be used to guide ML methods to facilitate systematic exploration, effectively automating the trial-and-error process [5–7]. Integrating chemical information into such ML models further enhances their predictive capabilities, transforming the process into a more informed and efficient endeavor. This chemical information can be obtained through detailed simulations or advanced characterization techniques. Beyond refining the methods themselves, chemometrics and ML also play critical roles in the optimization of the analytical methods [8] as well as in analyzing and interpreting the results of these methods [9].

Unfortunately, despite the availability of numerous publications and reviews highlighting the promise of machine learning in the polymer field [1, 4–7, 9], its practical application remains limited in several key areas.

In this perspective article, we examine the current applications and future potential of artificial intelligence (AI), particularly ML, in advancing polymer science and characterization. We first clarify the terms AI and ML and then review a select number of areas in which ML has been demonstrated to deliver on the great promise. The reviewed

Tijmen S. Bos t.s.bos@uva.nl

<sup>&</sup>lt;sup>1</sup> Analytical Chemistry Group, Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam, The Netherlands

<sup>&</sup>lt;sup>2</sup> Centre for Analytical Sciences Amsterdam (CASA), Amsterdam, The Netherlands

literature is then used to recalibrate expectations toward future outlook.

# Definitions of Machine Learning and Artificial Intelligence

It is essential to first clarify the term AI as it is often misapplied to any automated workflow or algorithm. AI specifically refers to algorithms or robots that are capable of mimicking and surpassing human capabilities by perceiving and interacting with an environment [10]. In the subset of AI known as ML, a common approach involves providing solutions based on a set of parameters either directly or iteratively. The outcomes are evaluated through one or more scores, and depending on the algorithm, it either seeks to maximize these scores or predict them based on parameter combinations. A relevant example is an ML algorithm designed to optimize specific or various properties of a polymer by iteratively refining the synthesis process based on feedback [11].

# Machine Learning to Aid Polymer Development

# **Simulation of Properties of Polymers**

An effective initial approach to identify which chemical properties influence the physical properties of a final polymer product is to utilize simulations. While polymer simulations often do not fully replicate real-world conditions, they provide valuable insights into identifying aspects of interest. For both simulation-derived and experimentally obtained data, it is crucial to describe structural information in an organized and standardized manner.

One framework that facilitates this standardization is Polydat, which allows for the recording of both structural data and characterized parameters [12]. Such efforts toward standardization can greatly benefit the polymer characterization community, enabling models to more readily integrate and utilize data from other researchers. Commonly, general polymer structures are reported using BigSMILES notation [13–16], an extension of the normal SMILES format that incorporates features specific to polymers, such as repeating units, branching, and end groups. However, BigSMILES representations are often too complex for direct use in the training of models. To address this, molecular descriptors can be employed to simplify the structural information [17]. This reduction in parameter space not only streamlines model development but also helps generalize the features responsible for the properties of interest. This standardization effort helps in the creation of a database that enables the research community to collaborate to build more advanced models [18, 19]. Once polymers are accurately described, ML models can be trained to predict their properties based on composition and/or structural information [11].

#### **Optimization of Properties**

ML proves highly effective for optimizing measurable properties and modifying synthesis conditions to enhance those and potentially other related properties [20, 21]. Notable examples of such applications include the design of polymer-based biomaterials [22], polymeric long-acting injectables [23], and orodispersible films used for drug delivery [24]. To perform these optimizations, it is first necessary to define the features of interest (polymer properties) and the adjustable variables [25]. The number of variables that can be adjusted simultaneously is often constrained as the search space expands exponentially with each additional variable. Features of interest can be selected either manually by the analyst or automatically using tools such as principal component analysis (PCA). Following the optimization process, interpreting the impact of various features is highly valuable. Eliminating features with minimal impact streamlines the workflow while the insights gained contribute to a deeper understanding of the underlying mechanisms [26].

#### **Polymer Synthesis and Discovery**

As described in the previous section, the synthesis conditions may be modified to achieve more desirable polymers. Some works have developed closed-loop automated workflows to achieve this while incorporating ML, flow chemistry synthesis, and automated chemical analysis. The use of flow chemistry for optimizing synthesis has been demonstrated for multiple applications, including optimizing the yield for photocatalysis [27], and the optimization of functionality of various polymers [28]. For polymer synthesis, a property of interest may be the monomer conversion where ideally 100% of the monomers are converted. To achieve this, a flow reactor was connected with nuclear magnetic resonance (NMR) [29]. Moreover, it can be coupled to size-exclusion chromatography (SEC) to assess the molar mass dispersity as a second parameter [30]. The closed-loop system automatically processes the SEC and NMR data to obtain the dispersity and monomer conversion. This data was fed into a Thompson sampling efficient multi-objective optimization (TS-EMO) model with every iteration to predict and subsequently identify the Pareto front for these objectives. It should be noted that this workflow can only identify the Pareto optimum within the user-defined reaction space (e.g., a maximum residence time of 20 min and a temperature range of 80-120 °C). While the previous synthetic routes were performed in flow, some reactions are better suited for batch processes. As such, a similar automated setup may be designed that enables batch operation modes [31]. It has, however, not yet been combined with online characterization and ML algorithms for the prediction of polymer properties.

Besides optimizing synthesis parameters, a similar system can be used to discover better-performing copolymers. Using a similar TS-EMO-based algorithm, conflicting optimization objectives, such as cost and yield, may be optimized using inline reversed-phase LC (RPLC) to determine yield for a single-step synthesis [32]. Automated flow synthesis has been used to discover <sup>19</sup>F magnetic resonance imaging agents by coupling it directly with <sup>19</sup>F NMR analysis [33]. Using a six-variable compositional space, 397 unique copolymer compositions were synthesized of which > 10 outperformed state-of-the-art materials, demonstrating such an approach to be efficient at tackling high-dimensional structure-property relationships that are otherwise difficult to model. Moreover, the discovery of new polymer designs for flame retardants was explored using a ML-assisted approach while manually synthesizing the proposed polymers [34].

# Analytical Methods Developed and Enhanced by Machine Learning

#### Chromatography

# Chromatographic Response Functions Are Needed to Drive ML Algorithms

There are many ML approaches developed already that are capable of aiding in the development of LC methods [8, 35, 36]. However, incorporating them in a closed-loop fashion remains challenging but has been demonstrated in a few examples using a Bayesian optimization algorithm [37–39]. The prime bottleneck appears to be the development of a chromatographic response function (CRF) that can guide the optimization process.

The above Bayesian optimization approaches were demonstrated using small molecules. Polymer characterization through LC often does not yield single peaks to be separated, but rather distributions. Defining a CRF that a ML algorithm can optimize is therefore more challenging. The simplest way to define such a function is by trying to achieve a resolution of at least 1.5 between neighboring peaks. While this may not be applicable to classical synthetic polymers, it has been investigated for oligonucleotides [40]. The authors demonstrated the use of a support vector regression (SVR) model to predict the resolution between impurities in 12and 16-mer oligonucleotide sequences for ion-pair reversedphase LC. This model may be used to aid in predicting suitable method conditions for arbitrary sequences [40].

Nevertheless, the development of specialized CRFs tailored to distribution analysis would be highly beneficial. There appear to be at least two potential strategies for such a CRF. [41]. The first strategy would aim to enhance resolution within a single distribution by stretching it as much as possible, providing more detailed insights into the resolved distribution. However, this approach has practical limitations as excessive stretching can result in a distribution that is too wide, leading to low signal intensity and poor detection at the outer edges of the distribution. The second strategy would focus on maximizing the separation between multiple distributions. To incorporate these strategies into CRFs and enable automated optimization for distribution separation, it is essential to accurately characterize the distributions. The simplest method for this involves characterizing distributions using their average moments, such as mean retention/elution time, asymmetry, and kurtosis.

#### Prediction of Polymer Solubility to Optimize Separations

An intriguing property of synthetic polymers is their ability, or inability, to dissolve in specific solvents. This characteristic is particularly critical in techniques such as liquid chromatography where the analyte must be soluble in at least the strong solvent used during the process. Accurately predicting the solubility of novel polymers can be challenging. However, modeling and predicting this behavior could provide significant advantages in both research and application contexts. ML approaches can be employed to estimate the Flory-Huggins interaction parameter of a polymer-solvent mixture [42-44], which provides valuable insights into solubility. Alternatively, neural networks can be trained using large databases that include solvent compatibility information for various polymers. By leveraging polymer structural data, these models can predict suitable solvents with a high degree of accuracy [45, 46].

While these solubility models have not yet been applied to the optimization of chromatographic separations, they do offer a potential solution for automated method optimization. Common chromatographic modes for synthetic polymers rely on polymer precipitation and subsequently redissolving for elution. This is performed using either a solvent gradient, which is referred to as gradient polymer-elution chromatography (GPEC), or a temperature gradient, referred to as temperature gradient interaction chromatography (TGIC). Moreover, to perform LC under critical conditions, a combination of a "good" and "bad" solvent is required in a specific ratio to obtain a mass-independent separation condition. While these critical solvent conditions are known for a handful of polymers, it would be a great advantage if such conditions could be predicted by solubility models.

#### **Detection and Identification Methods**

#### Detection

While significant progress has been made in applying ML to method development on the chromatographic side, advancements in detection techniques and the analysis of resulting polymer data have been comparatively limited. Frequently, dedicated algorithms based on expert knowledge are employed to extract specific information from the data [47–49]. For ML to be effectively utilized in this domain, it is crucial to generalize problems across multiple polymer types. Nevertheless, ML has demonstrated its utility in efficiently solving complex problems, particularly in fitting intricate models to challenging datasets. An example of this is the determination of block length distributions in copolymers based on fragment data [50].

#### Identification

Machine learning can be highly advantageous for processing raw data and translating it into more informative insights. A common application is the identification of classes within datasets, typically achieved by training a model on labeled data where the class distinctions are known. For instance, ML has been employed to identify microplastics [51]. In one study, a random forest model was trained to distinguish between polymethyl methacrylate (PMMA), polystyrene (PS), polytetrafluoroethylene (PTFE), polyvinylchloride (PVC), and polyethylene (PE) based on Raman spectra obtained from environmental samples [52]. Similarly, leveraging infrared (IR) spectroscopy data, classification techniques have been used to identify plasticizers in PVC [53] and nylon particles [54].

The same principles can be extended to mass spectrometry (MS) data, including imaging MS [55]. While these identification methods typically determine the most likely class, they often provide limited insights into the probability of the classification. Logistic regression is frequently employed for such tasks, and when threshold criteria are omitted, it can also be used to report probabilities for various classes. This approach has been applied, for instance, to identify monomers leaching from dental composites [56]. ML has also been employed to discriminate between virgin and recycled poly(ethylene terephthalate) (PET) based on data from headspace comprehensive two-dimensional gas chromatography coupled with mass spectrometry [57].

Beyond direct identification, ML is also applied to predict polymer characteristics such as their size. These predictions can assist in identification processes, particularly with techniques like ion mobility spectrometry where ML is used to predict the collision cross-section of polymers. Such predictions help narrow the search window for potential candidates, enhancing the efficiency and accuracy of the identification process [58].

### **Future Perspectives**

ML is increasingly applied across various aspects of polymer development and characterization, accelerating research and expanding capacity. While ML often enhances a single component of the workflow, its potential extends to multiple stages of the process. Although ML has demonstrated success in optimizing synthesis procedures to improve polymer properties and classifying polymers using established methods, greater attention should be directed toward automating method development for characterization techniques as well as improving data processing and interpretation. This shows that the readiness of different aspects of polymer science to benefit from ML varies significantly. These varying levels of readiness are depicted in Fig. 1. Addressing the lessdeveloped areas is essential for achieving a fully versatile and integrated workflow. The primary areas with significant potential for advancement are detection techniques, chromatography, and the analysis of the resulting data. Applying ML in these fields to optimize methods offers the potential to greatly enhance sensitivity and separation performance. Furthermore, ML-based models for data interpretation can uncover novel insights into structure-property relationships that were previously inaccessible.

Advancements in ML are expected to accelerate the optimization process, thereby reducing the number of steps and measurements required in experimental workflows. Another promising avenue is the integration of ML with expert knowledge of fundamental principles. This can be achieved by supplementing experimental data with data from simulations. A recent example of an algorithm capable of handling this approach is multi-task Bayesian optimization [37]. These algorithms can identify and correct biases in simulations based on experimental data, enabling efficient optimization of more complex problems.

A distinctly different, yet rapidly advancing class, of algorithms is large language models (LLMs). LLMs are fundamentally trained to generate and predict patterns based on vast amounts of data. They "mimic" understanding rather than develop a mechanistic model of the process being optimized. While LLMs can suggest optimizations based on learned patterns, they do not explicitly model uncertainty or actively search for the optimal solution in a structured way. While LLMs are unlikely to contribute directly to optimization or characterization processes, since they do not inherently understand data but rather mimic understanding, they hold potential in facilitating more intuitive interactions with laboratory equipment and specialized data processing software. Additionally, they could play a valuable role in streamlining



**Fig. 1** Flowchart illustrating the various aspects of polymer development that can be enhanced by optimizing functional properties through polymer characterization. The colors represent the current

technology readiness level (TRL) of machine learning applicability in each aspect of the application for polymers as defined by the European Commission in the Horizon [59]

reporting and documentation processes, enhancing accessibility and efficiency in polymer research.

We envision a future where these advancements empower researchers to dedicate more time to addressing scientific questions and interpreting results while also enhancing and streamlining method development and synthesis processes. While this level of automation may require years or decades to achieve, the eventual integration of laboratory processes could lead to the creation of interfaces capable of recommending experiments, within the laboratory's capabilities, to effectively answer specific research questions.

Acknowledgements This research is part of the PARADISE project (ENPPS.TA.019.001) and received funding from the Dutch Research Council (NWO) in the framework of the Science PPP Fund for the top sectors and from the Ministry of Economic Affairs of the Netherlands in the framework of the "PPS Toeslagregeling".

Author Contributions Rick S. van den Hurk: Conceptualization, Writing—original draft. Bob W.J. Pirok: Conceptualization, Funding acquisition, Writing—review & editing. Tijmen S. Bos: Conceptualization, Writing—original draft.

Funding Nederlandse Organisatie voor Wetenschappelijk Onderzoek, ENPPS.TA.019.001.

**Data Availability** No datasets were generated or analysed during the current study.

#### Declarations

Conflict of interest The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits 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/4.0/.

# References

- Vriza A, Chan H, Xu J (2023) Self-driving laboratory for polymer electronics. Chem Mater 35:3046–3056. https://doi.org/10.1021/ acs.chemmater.2c03593
- Lee J, Mulay P, Tamasi MJ et al (2023) A fully automated platform for photoinitiated RAFT polymerization. Digit Discov 2:219–233. https://doi.org/10.1039/D2DD00100D
- Gongora AE, Xu B, Perry W et al (2024) A Bayesian experimental autonomous researcher for mechanical design. Sci Adv 6:eaaz1708. https://doi.org/10.1126/sciadv.aaz1708
- Beaucage PA, Sutherland DR, Martin TB (2024) Automation and machine learning for accelerated polymer characterization and development: past, potential, and a path forward. Macromolecules 57:8661–8670. https://doi.org/10.1021/acs.macromol.4c01410
- Martin TB, Audus DJ (2023) Emerging trends in machine learning: a polymer perspective. ACS Polym Au 3:239–258. https:// doi.org/10.1021/acspolymersau.2c00053

- Jayaraman A, Olsen B (2024) Convergence of artificial intelligence, machine learning, cheminformatics, and polymer science in macromolecules. Macromolecules 57:7685–7688. https://doi. org/10.1021/acs.macromol.4c01704
- Cencer MM, Moore JS, Assary RS (2022) Machine learning for polymeric materials: an introduction. Polym Int 71:537–542. https://doi.org/10.1002/pi.6345
- van Henten GB, Bos TS, Pirok B (2023) Approaches to accelerate liquid chromatography method development in the laboratory using chemometrics and machine learning. LCGC Europe 36:202–209. https://doi.org/10.56530/lcgc.eu.rh7676j5
- 9. Houhou R, Bocklitz T (2021) Trends in artificial intelligence, machine learning, and chemometrics applied to chemical data. Anal Sci Adv 2:128–141. https://doi.org/10.1002/ansa.202000162
- 10. Russell S, Norvig P (2020) Artificial intelligence: a modern approach, 4th edn. Pearson
- 11. Queen O, McCarver GA, Thatigotla S et al (2023) Polymer graph neural networks for multitask property learning. NPJ Comput Mater 9:90. https://doi.org/10.1038/s41524-023-01034-3
- Lin T-S, Rebello NJ, Beech HK et al (2021) PolyDAT: a generic data schema for polymer characterization. J Chem Inf Model 61:1150–1163. https://doi.org/10.1021/acs.jcim.1c00028
- Lin T-S, Coley CW, Mochigase H et al (2019) BigSMILES: a structurally-based line notation for describing macromolecules. ACS Cent Sci 5:1523–1531. https://doi.org/10.1021/acscentsci. 9b00476
- Lin T-S, Rebello NJ, Lee G-H et al (2022) Canonicalizing BigSMILES for polymers with defined backbones. ACS Polym Au 2:486–500. https://doi.org/10.1021/acspolymersau.2c00009
- Zou W, Martell Monterroza A, Yao Y et al (2022) Extending BigSMILES to non-covalent bonds in supramolecular polymer assemblies. Chem Sci 13:12045–12055. https://doi.org/10.1039/ D2SC02257E
- Deagen ME, Dalle-Cort B, Rebello NJ et al (2024) Machine translation between BigSMILES line notation and chemical structure diagrams. Macromolecules 57:42–53. https://doi.org/10.1021/acs. macromol.3c01378
- Liu AL, Venkatesh R, McBride M et al (2020) Small data machine learning: classification and prediction of poly(ethylene terephthalate) stabilizers using molecular descriptors. ACS Appl Polym Mater 2:5592–5601. https://doi.org/10.1021/acsapm.0c00921
- Kim C, Chandrasekaran A, Huan TD et al (2018) Polymer genome: a data-powered polymer informatics platform for property predictions. J Phys Chem C 122:17575–17585. https://doi. org/10.1021/acs.jpcc.8b02913
- Sha W, Li Y, Tang S et al (2021) Machine learning in polymer informatics. InfoMat 3:353–361. https://doi.org/10.1002/inf2. 12167
- Upadhya R, Kosuri S, Tamasi M et al (2021) Automation and data-driven design of polymer therapeutics. Adv Drug Deliv Rev 171:1–28. https://doi.org/10.1016/j.addr.2020.11.009
- 21. Sharma A, Mukhopadhyay T, Rangappa SM et al (2022) Advances in computational intelligence of polymer composite materials: machine learning assisted modeling, analysis and design. Arch Comput Methods Eng 29:3341–3385. https://doi.org/10.1007/ s11831-021-09700-9
- Patel RA, Webb MA (2024) Data-driven design of polymer-based biomaterials: high-throughput simulation, experimentation, and machine learning. ACS Appl Bio Mater 7:510–527. https://doi. org/10.1021/acsabm.2c00962
- Bannigan P, Bao Z, Hickman RJ et al (2023) Machine learning models to accelerate the design of polymeric long-acting injectables. Nat Commun 14:35. https://doi.org/10.1038/ s41467-022-35343-w
- 24. O'reilly CS, Elbadawi M, Desai N et al (2021) Machine learning and machine vision accelerate 3d printed orodispersible film

development. Pharmaceutics. https://doi.org/10.3390/pharmaceut ics13122187

- Day EC, Chittari SS, Bogen MP, Knight AS (2023) Navigating the expansive landscapes of soft materials: a user guide for highthroughput workflows. ACS Polym Au 3:406–427. https://doi.org/ 10.1021/acspolymersau.3c00025
- Munir N, McMorrow R, Mulrennan K et al (2023) Interpretable machine learning methods for monitoring polymer degradation in extrusion of polylactic acid. Polymers (Basel). https://doi.org/10. 3390/polym15173566
- Slattery A, Wen Z, Tenblad P et al (2025) Automated self-optimization, intensification, and scale-up of photocatalysis in flow. Science 383:eadj1817. https://doi.org/10.1126/science.adj1817
- Yu L, Chen B, Li Z et al (2024) Computer-aided automated flow chemical synthesis of polymers. Giant 18:100252. https://doi.org/ 10.1016/j.giant.2024.100252
- Rubens M, Van Herck J, Junkers T (2019) Automated polymer synthesis platform for integrated conversion targeting based on inline benchtop NMR. ACS Macro Lett 8:1437–1441. https:// doi.org/10.1021/acsmacrolett.9b00767
- Knox ST, Parkinson SJ, Wilding CYP et al (2022) Autonomous polymer synthesis delivered by multi-objective closed-loop optimisation. Polym Chem 13:1576–1585. https://doi.org/10. 1039/D2PY00040G
- Jafari VF, Mossayebi Z, Allison-Logan S et al (2023) The power of automation in polymer chemistry: precision synthesis of multiblock copolymers with block sequence control. Chem Eur J 29:e202301767. https://doi.org/10.1002/chem.202301767
- Jeraal MI, Sung S, Lapkin AA (2021) A machine learningenabled autonomous flow chemistry platform for process optimization of multiple reaction metrics. Chem-Methods 1:71–77. https://doi.org/10.1002/cmtd.202000044
- Reis M, Gusev F, Taylor NG et al (2021) Machine-learningguided discovery of 19F MRI agents enabled by automated copolymer synthesis. J Am Chem Soc 143:17677–17689. https://doi.org/10.1021/jacs.1c08181
- Wang R, Fu T, Yang Y-J et al (2024) Scientific discovery framework accelerating advanced polymeric materials design. Research 7:0406. https://doi.org/10.34133/research.0406
- Bosten E, Pardon M, Chen K et al (2024) Assisted active learning for model-based method development in liquid chromatography. Anal Chem 96:13699–13709. https://doi.org/10.1021/acs. analchem.4c02700
- Kensert A, Libin P, Desmet G, Cabooter D (2024) Deep reinforcement learning for the direct optimization of gradient separations in liquid chromatography. J Chromatogr A 1720:464768. https://doi.org/10.1016/j.chroma.2024.464768
- Boelrijk J, Molenaar SRA, Bos TS et al (2024) Enhancing LC×LC separations through multi-task Bayesian optimization. J Chromatogr A 1726:464941. https://doi.org/10.1016/j.chroma. 2024.464941
- Bos TS, Boelrijk J, Molenaar SRA et al (2022) Chemometric strategies for fully automated interpretive method development in liquid chromatography. Anal Chem 94:16060–16068. https:// doi.org/10.1021/acs.analchem.2c03160
- Boelrijk J, Ensing B, Forré P, Pirok BWJ (2023) Closed-loop automatic gradient design for liquid chromatography using Bayesian optimization. Anal Chim Acta 1242:340789. https:// doi.org/10.1016/j.aca.2023.340789
- Enmark M, Häggström J, Samuelsson J, Fornstedt T (2022) Building machine-learning-based models for retention time and resolution predictions in ion pair chromatography of oligonucleotides. J Chromatogr A 1671:462999. https://doi.org/10. 1016/j.chroma.2022.462999
- Bos TS (2023) Chemometric tools for automated methoddevelopment and data interpretation in liquid chromatography.

PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam. https://doi.org/10.5463/thesis.464

- 42. Nistane J, Chen L, Lee Y et al (2022) Estimation of the Flory– Huggins interaction parameter of polymer-solvent mixtures using machine learning. MRS Commun 12:1096–1102. https:// doi.org/10.1557/s43579-022-00237-x
- Liu T-L, Liu L-Y, Ding F, Li Y-Q (2022) A machine learning study of polymer–solvent interactions. Chin J Polym Sci 40:834–842. https://doi.org/10.1007/s10118-022-2716-2
- 44. Aoki Y, Wu S, Tsurimoto T et al (2023) Multitask machine learning to predict polymer–solvent miscibility using Flory– Huggins interaction parameters. Macromolecules 56:5446– 5456. https://doi.org/10.1021/acs.macromol.2c02600
- 45. Chandrasekaran A, Kim C, Venkatram S, Ramprasad R (2020) A deep learning solvent-selection paradigm powered by a massive solvent/nonsolvent database for polymers. Macromolecules 53:4764–4769. https://doi.org/10.1021/acs.macromol.0c00251
- 46. Peacock CJ, Lamont C, Sheen DA et al (2021) Predicting the mixing behavior of aqueous solutions using a machine learning framework. ACS Appl Mater Interfaces 13:11449–11460. https:// doi.org/10.1021/acsami.0c21036
- Molenaar SRA, van de Put B, Desport JS et al (2022) Automated feature mining for two-dimensional liquid chromatography applied to polymers enabled by mass remainder analysis. Anal Chem 94:5599– 5607. https://doi.org/10.1021/acs.analchem.1c05336
- Desport JS, Frache G, Patiny L (2020) MSPolyCalc: a web-based app for polymer mass spectrometry data interpretation—the case study of a pharmaceutical excipient. Rapid Commun Mass Spectrom 34:e8652. https://doi.org/10.1002/rcm.8652
- Vlnieska V, Khanda A, Gilshtein E et al (2024) Polypy: a framework to interpret polymer properties from mass spectroscopy data. Polymers (Basel). https://doi.org/10.3390/polym16131771
- van den Hurk RS, Mengerink Y, Peters RAH et al (2024) Introducing an algorithm to accurately determine copolymer block-length distributions. Anal Chim Acta 343990. https://doi.org/10.1016/j. aca.2025.343990
- Lin J, Liu H, Zhang J (2022) Recent advances in the application of machine learning methods to improve identification of the microplastics in environment. Chemosphere 307:136092. https:// doi.org/10.1016/j.chemosphere.2022.136092
- Xie L, Luo S, Liu Y et al (2023) Automatic identification of individual nanoplastics by raman spectroscopy based on machine learning. Environ Sci Technol 57:18203–18214. https://doi.org/ 10.1021/acs.est.3c03210
- Rijavec T, Ribar D, Markelj J et al (2022) Machine learningassisted non-destructive plasticizer identification and quantification in historical PVC objects based on IR spectroscopy. Sci Rep 12:5017. https://doi.org/10.1038/s41598-022-08862-1
- Yang C, Xie J, Gowen A, Xu J-L (2024) Machine learning driven methodology for enhanced nylon microplastic detection and characterization. Sci Rep 14:3464. https://doi.org/10.1038/ s41598-024-54003-1
- Jetybayeva A, Borodinov N, Ievlev AV et al (2023) A review on recent machine learning applications for imaging mass spectrometry studies. J Appl Phys 133:020702. https://doi.org/10.1063/5.0100948
- Chen C, Mondal K, Vervliet P et al (2023) Logistic regression analysis of LC-MS/MS data of monomers eluted from aged dental composites: a supervised machine-learning approach. Anal Chem 95:5205–5213. https://doi.org/10.1021/acs.analchem.2c04362
- Li H, Wu X, Wu S et al (2022) Machine learning directed discrimination of virgin and recycled poly(ethylene terephthalate) based on non-targeted analysis of volatile organic compounds. J Hazard Mater 436:129116. https://doi.org/10.1016/j.jhazmat. 2022.129116
- 58. Foster M, Rainey M, Watson C et al (2022) Uncovering PFAS and other xenobiotics in the dark metabolome using ion mobility

spectrometry, mass defect analysis, and machine learning. Environ Sci Technol 56:9133–9143. https://doi.org/10.1021/acs.est.2c00201

 European Commission: Horizon 2020 – Work Programme 2014– 2015, Annex G. https://ec.europa.eu/research/participants/data/ref/ h2020/wp/2014\_2015/annexes/h2020-wp1415-annex-g-trl\_en.pdf

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.





Rick S. van den Hurk finalized his Ph.D. earlier this year at the University of Amsterdam. During this time, he developed modulation interfaces for the application of 2D liquid chromatography and applied these to characterize nitrocellulose-based explosive materials. He also developed computational polymer analytics. Van der Hurk recently won the poster prize at the Multidimensional Chromatography Workshop in Belgium earlier this vear.

Bob W. J. Pirok obtained his Ph.D. cum laude in 2019 in Amsterdam and is currently an associate professor at the University of Amsterdam. His research group focuses on the application of chemometrics, including machine learning, to analytical chemistry with a special interest in method development and data analysis for one and two-dimensional chromatography. Pirok is a visiting research professor at Gustavus Adolphus College in the group of Prof. Dwight R. Stoll. He received the HTC Innovation Award in 2024 for his work

automation of method development in liquid chromatography. Pirok is also the author of the educational textbook Analytical Separation Science, which is slated to be published in May 2025.



Tijmen S. Bos earned his Ph.D. in 2023 at the Vrije Universiteit Amsterdam. He currently is a postdoctoral researcher at the University of Amsterdam, specializing in the automated development of methods for chromatographic systems and advanced data analysis. His research is particularly centered on the analysis of macromolecules such as polymers, striving to enhance separation capabilities and develop novel algorithms to extract maximum insights from analytical data.