



(51) International Patent Classification:

G06Q 10/0631 (2023.01) G06N 3/02 (2006.01)

G06Q 50/04 (2012.01)

(21) International Application Number:

PCT/EP2024/085680

(22) International Filing Date:

11 December 2024 (11.12.2024)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

23216804.7 14 December 2023 (14.12.2023) EP

24155217.3 01 February 2024 (01.02.2024) EP

24155823.8 05 February 2024 (05.02.2024) EP

(71) Applicant: BASF SE [DE/DE]; Carl-Bosch-Strasse 38, 67056 Ludwigshafen am Rhein (DE).

(72) Inventors: SALENTIN, Sebastian; Speyerer Straße 2, 67117 Limburgerhof (DE). EN-NASER, Gerhard; Carl-Bosch-Straße 38, 67056 Ludwigshafen am Rhein (DE). LI, Yueling; Pfalzgrafenstraße 1, 67061 Ludwigshafen am Rhein (DE). MARTSCHAT, Sebastian Hermann; Pfalzgrafenstraße 1, 67061 Ludwigshafen am Rhein (DE). MATHEA, Miriam; Carl-Bosch-Straße 38, 67056 Ludwigshafen am Rhein (DE). SAMIEE, Soheila; 5025 Creekbank Rd, Mississauga, Ontario L4W 0B6 (CA). SANDFORT, Frederik; Carl-Bosch-Straße 38, 67056 Ludwigshafen am Rhein (DE). STORK, Conrad; Carl-Bosch-Straße 38, 67056 Ludwigshafen am Rhein (DE).

(74) Agent: BASF IP ASSOCIATION; BASF SE GBI - Z078, 67056 Ludwigshafen (DE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG,

KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE

(54) Title: AGENT SELECTION SERVICE FOR CHEMICAL INDUSTRY

(57) Abstract: A method for providing a target chemical product associated with chemical product data characterizing the target chemical product and/or one or more properties of the chemical product, the method comprising: providing a request for providing the target chemical product, wherein the request and is associated with an indication on the target chemical product, providing functional specification data related to one or more functions of one or more operating engine(s) for providing the chemical product data, providing a selection task instruction related to the request and the functional specification data to a selection model for generating operating input data, wherein the selection model is configured to select at least one operating engine(s) according to the request and the functional specification data, wherein the operating input data is indicative of the at least one selected operating engine and wherein the operating input data is suitable for triggering the selected operating engine, providing the operating input data to the at least one selected operating engine for generating the chemical product data, wherein the selected operating engine is configured to provide at least a part of the chemical product data in response to providing the operating input data, providing the generated chemical product data for providing the target chemical product with the one or more target properties.

AGENT SELECTION SERVICE FOR CHEMICAL INDUSTRY

TECHNICAL FIELD

5 The disclosure relates to chemical production, in particular tailored to customer needs, and to use of chemical product data, use of a selection task instruction, use of a structure model, method for producing a chemical product with one or more target properties, a method for producing and/or processing a target chemical product associated with chemical product data characterizing the target chemical product and/or one or more properties of the chemical product, method for generating chemical product data characterizing a chemical product with one or
10 more target properties, apparatus in particular for producing and/or processing a target chemical product associated with chemical product data characterizing the target chemical product and/or one or more properties of the chemical product.

TECHNICAL BACKGROUND

15 Chemical production networks are complex production systems for producing hundreds of distinct chemical products with varying properties. Hence, providing chemical products with tailored properties is challenging.

SUMMARY

20 In an aspect this disclosure relates to a method, in particular computer-implemented method, for generating chemical product data characterizing a chemical product with one or more target properties, the method comprising:

obtaining, in particular receiving, preferably via an interface such as user interface, a request for providing
25 the chemical product with the one or more target properties, wherein the request includes unstructured data and is associated with one or more target properties of the chemical product,
obtaining, in particular receiving, preferably via an interface such as user interface, functional specification data related to one or more functions of one or more operating engine(s) for providing the chemical product with the one or more target properties,
30 providing one or more input data structure related to input data suitable for being provided to the one or more operating engine(s),
providing, in particular by a processing device and/or an interface, a task instruction including unstructured data related to the request, the one or more input data structure(s) and the functional specification to one or more data-driven model(s), wherein the one or more data-driven model(s) are configured to generate
35 operating input data for one or more selected operating engine(s) in relation to the provided one or more

target properties, wherein the operating input data includes structured data for triggering the selected operating engine,

providing, in particular by a processing device, the operating input data to the at least one selected operating engine for generating the chemical product data, wherein the operating engine is configured to generate at least a part of the chemical product data in response to providing the operating input data, providing, in particular via an interface such as a user interface, the generated chemical product data for producing and/or processing the chemical product with the one or more target properties, optionally producing and/or processing the chemical product with the one or more target properties.

10 In another aspect, it relates to a method for producing and/or processing a target chemical product associated with chemical product data characterizing the target chemical product and/or one or more properties of the target chemical product, the method comprising:

obtaining, in particular receiving, a request, in particular via an interface such as a user interface, for providing the chemical product data, wherein the request is associated with an indication on the target chemical product,

obtaining, in particular receiving, in particular by a processing device such as a processor, functional specification data related to one or more functions of two or more operating engine(s) for providing the chemical product data,

providing, in particular by a processing device such as a processor, a selection task instruction related to the request and the functional specification data to a selection model for generating operating input data, wherein the selection model is configured to select at least one operating engine(s) according to the request and the functional specification data, wherein the operating input data is indicative of the at least one selected operating engine and wherein the operating input data is suitable for triggering the selected operating engine,

providing, in particular by a processing device such as a processor, the operating input data to the at least one selected operating engine for generating the chemical product data, wherein the selected operating engine is configured to provide at least a part of the chemical product data in response to providing the operating input data to the at least one selected operating engine,

providing, in particular by a processing device such as a processor, the generated chemical product data for producing and/or processing the target chemical product with the one or more target properties.

In another aspect, it relates to a method for producing and/or processing a target chemical product associated with chemical product data characterizing the target chemical product and/or one or more properties of the chemical product, the method comprising:

providing a request for providing the chemical product data, wherein the request includes unstructured data and is associated with an indication on the target chemical product,

providing functional specification data related to one or more functions of two or more operating engine(s) for providing the chemical product data,

5 providing a selection task instruction including unstructured data related to the request and the functional specification data to a selection model for generating operating input data, wherein the selection model is configured to select at least one operating engine(s) according to the request and the functional specification data, wherein the operating input data is indicative of the at least one selected operating engine and wherein the operating input data is suitable for triggering the selected operating engine,

10 providing the operating input data to the at least one selected operating engine for generating the chemical product data, wherein the selected operating engine is configured to provide at least a part of the chemical product data in response to providing the operating input data to the at least one selected operating engine, providing the generated chemical product data for producing and/or processing the target chemical product with the one or more target properties.

15 In another aspect, it relates to an apparatus comprising: a processor; and a memory storing instructions that, when executed by the processor, configure the apparatus to perform any one of the methods described herein.

In another aspect, it relates to use of chemical product data as obtained by any one of the methods described
20 herein for producing and/or processing a chemical product.

In another aspect, it relates to use of a selection task instruction according to any one of the methods described herein for processing a request for providing the chemical product with the one or more target properties according to any one of the methods described herein.

25 In another aspect, it relates to use of selection model as described herein for providing chemical product data for producing and/or processing a chemical product.

In another aspect, it relates to a method for producing a chemical product with one or more target properties, the
30 method comprising:

providing a request for producing the chemical product with the one or more target properties, wherein the request includes unstructured data and is associated with one or more target properties of the chemical product,

35 providing functional specification data related to one or more functions of one or more operating engine(s) for providing the chemical product with the one or more target properties,

providing one or more input data structure related to input data suitable for being provided to the one or more operating engine(s),
providing a task instruction including unstructured data related to the request, the one or more input data structure(s) and the functional specification to one or more data-driven model(s), wherein the one or more
5 data-driven model(s) are configured to generate operating input data for one or more selected operating engine(s) in relation to the provided one or more target properties, wherein the operating input data includes structured data for triggering the selected operating engine,
providing the operating input data to the at least one selected operating engine for generating the chemical product data characterizing the chemical product with the one or more target properties, wherein the
10 operating engine is configured to generate at least a part of the chemical product data in response to providing the operating input data,
producing the chemical product based on the chemical product data.

EMBODIMENTS

15 Any disclosure, embodiments and examples described herein relate to the methods, the systems, apparatuses, chemical products and computer elements lined out above and below. Advantageously, the benefits provided by any of the embodiments and examples equally apply to all other embodiments and examples.

20 In the following, terminology as used herein and/or the technical field of the present disclosure will be outlined by ways of definitions and/or examples. Where examples are given, it is to be understood that the present disclosure is not limited to said examples.

These and other objects, which become apparent upon reading the following description, are solved by the subject
25 matters of the independent claims. The dependent claims refer to embodiments of the disclosure.

Chemical products are starting materials for a plurality of different end products. As a consequence, chemical products have to provide a variety of changing properties tailored to the intended end product. The production of chemical products starts with raw materials that are processed via one or more processing steps including for
30 example chemical reactions conducted in reactors and purification steps. Typically, chemical products are obtained from two or more chemical reactions changing the chemical structure of the reactants and thus, changing the properties of the chemical product. For example, a liquid such as monoethylenglycol and a solid such as terephthalic acid can be converted to yield polyester. Polyester is a functional polymer with distinct properties depending on the educts and reaction conditions. This may lead to a polyester with different properties.
35 Accordingly, the chemical reaction to produce polyester needs to be tailored to the desired properties of the

polyester. The polyester can be applied in a variety of field such as clothing or packaging. Hence, the challenge constitutes in serving hundreds of customers with thousands of distinct chemical products obtained from a chemical production network with a plurality of production steps to result in distinct and tailored properties of the chemical products. The properties of chemical products highly depend on its chemical structure. Even a small change in the orientation of a subgroup of a molecule with hundreds of atoms results in a distinct chemical property. Therefore, a relation between a chemical structure of the chemical product and the properties of the chemical product is complex and challenging to control.

Processing a request for providing the chemical product with the one or more target properties allows for obtaining target chemical products in response to receiving unstructured requests. This is particularly advantageous as the indication on the target chemical products come in a plurality of different formats. For example, ethene may be the IUPAC name, whereas the same chemical compound may be associated with the trivial name ethylene. Further, companies sell chemical products under their established product name. Providing the request for selecting the part of operation instructions associated with an operation related to the request enables efficient and robust determination of chemical product data for producing target chemical products. Hence, reliable processing and/or production of chemical products associated with a target product.

In an embodiment, chemical product data may be associated with a digital representation of the chemical structure of the chemical product with the one or more target properties. In particular, the chemical product data may include the digital representation of the chemical structure of the chemical product with the one or more target properties. Further, the chemical product data may include unstructured data such as string data. The chemical product data may include machine instructions for providing the chemical product with the one or more target properties. The machine instructions may be provided for producing and/or processing of the chemical product with the one or more target properties. In an embodiment, the chemical product data may be associated with one or more properties of the target chemical product, in particular one or more properties as requested by the request for providing the chemical product data. Preferably, the chemical product data includes the one or more properties of the target chemical product, in particular one or more properties as requested by the request for providing the chemical product data.

In an embodiment, the indication on the target chemical product may be data indicative of the chemical product. The indication on the target chemical product may be related to the chemical product and/or the properties of the chemical product. The indication on the target chemical product may identify the chemical product. The indication on the target chemical product may be related to a chemical structure of the chemical product. Hence, the indication on the target chemical product may comprise a digital representation of the chemical structure of the chemical product.

In an embodiment, digital representation of the chemical structure may refer to a machine-readable and/or a machine-interpretable representation of the chemical structure of a chemical product. The digital representation of the chemical structure may be indicative of the chemical structure of the chemical product. The chemical structure may specify one or more atoms associated with a chemical product. Specifying the one or more atoms may refer to specifying one or more elements associated with the one or more atoms. Further, the digital representation of the chemical structure may be indicative of a relation between the one or more atoms, preferably an interaction between the one or more atoms, most preferably one or more bonds between the one or more atoms. Additionally or alternatively, digital representation of the chemical structure may be indicative of an arrangement of the one or more atoms in space, preferably in relation to a predefined point and/or to at least one atom of the one or more atoms. The digital representation of the chemical structure may include string data, in particular indicative of the one or more elements associated with the one or more atoms, and/or numerical data, in particular indicative of a relation between the one or more atoms.

In an embodiment, functional specification data may be related to one or more functions of the one or more operating engine(s). The functional specification data may comprise a functional specification associated with the one or more operating engine(s). The functional specification data may be indicative of a processing of operating input data by the one or more operating engine(s), in particular to operating output data. The functional specification associated with the one or more operating engine(s) may be indicative of one or more functions associated with the one or more operating engine(s), in particular the one or more functions carried out by the one or more operating engine(s). The one or more functions may include at least one function associated with providing the chemical product with the one or more target properties. The at least one function may configure the one or more operating engine(s), in particular the selected operating engine, to provide the chemical product with the one or more target properties. Further, the functional specification may be indicative of input data and/or output data associated with the one or more operating engine(s). The output data associated with the operating engine may be operating output data. The chemical product data may comprise at least a part of the operating output data. The functional specification data may include unstructured data.

In an embodiment, the one or more input data structure(s) may be indicative of one or more input data formats suitable for being provided to the one or more operating engine(s). The one or more input data structure(s) may refer to an arrangement of input data, in particular input data to the one or more operating engine(s). Operating input data may be input data to the one or more operating engine(s). The input data structure may be indicative of an arrangement of the one or more target properties, optionally a target type of chemical product and/or a target field of application associated with the chemical product. Operating input data associated with the one or more

input data structure(s) may be provided to the one or more operating engine(s), in particular the at least one selected operating engine.

5 In an embodiment, operating input data may be provided to the one or more operating engine(s), in particular the selected operating engine. The operating input data may comprise structured data. The operating input data may be associated with, in particular include, the indication on the target chemical product. The operating input data may be derived from and/or may depend on the indication on the target chemical product. The operating input data may be associated with a data structure related to selected operating engine. The one or more operating engine(s) may be configured to receive the operating input data and generating at least a part of the chemical
10 product data, in particular the operating output data, from the operating input data. The operating input data may be associated with a digital representation of the chemical structure of the chemical product, in particular may include a digital representation of the chemical structure of the chemical product. The operating input data may be derived from and/or may depend on the one or more target properties. The operating output data may be associated with, in particular relate to, more preferably comprise, at least a part of the chemical product data.

15 In an embodiment, the request for providing the chemical product with the one or more target properties may include unstructured data. The request may be associated with the one or more target properties of the chemical product. The request may include one or more target properties. Further, the request may include user instructions for providing the chemical product with the one or more target properties. The request and/or the user
20 instructions may include string data. The user instructions may be indicative of a function to be carried out by the selected operating engine. The request and/or the target properties may include numerical data. The request may further include an indication on the target chemical product. The indication on the target chemical product may be suitable for selecting a subgroup of chemical products. The indication of the chemical product may include a target type of the chemical product, a target field of application of the chemical product and/or a quantity
25 associated with the chemical product. The request may be provided and/or received via a user interface. The indication on the target chemical product and/or the operating input data may include a digital representation of the target chemical product. The operating input data may include at least a part of the request, in particular the indication on the target chemical product.

30 In an embodiment, the one or more target properties may be one or more properties required for processing the chemical product, in particular to an end product. The one or more target properties may include one or more chemical properties, one or more physical properties, one or more environmental attributes and/or one or more biological properties. The target property may include a user-specific target property. The user-specific target property may be provided and/or received via a user interface. The target property may be provided by a chemical

product processing facility. The chemical product processing facility may trigger the receiving to the request associated with the one or more target properties.

In an embodiment, environmental attribute may comprise at least one of emission data of the chemical product, recycle content of the chemical product, bio-based content of the chemical product, renewable content of the chemical product, chemical product declaration data, chemical product safety data or a combination thereof. Emission data may comprise any data related to environmental footprint. The environmental footprint may refer to an entity and its associated environmental footprint. The environmental footprint may be entity specific. For instance, the environmental footprint may relate to a chemical product, a company, a process such as a manufacturing process, a raw material or basic substance, a chemical product or material, a component, a component assembly, an end product, combinations thereof or additional entity-specific relations. Emission data may include data relating to carbon footprint of a chemical product. Emission data may include data relating to greenhouse gas emissions e.g. released in production of the chemical product. Emission data may include data related to greenhouse gas emissions. Greenhouse gas emissions may include emissions such as carbon dioxide (CO₂) emission, methane (CH₄) emission, nitrous oxide (N₂O) emission, hydrofluorocarbons (HFCs) emission, perfluorocarbons (PFCs) emission, sulphurhexafluoride (SF₆) emission, nitrogen trifluoride (NF₃) emission, combinations thereof and additional emissions. Emission data may include data related to greenhouse gas emissions of an entities or companies own operations (production, power plants and waste incineration). Scope 2 comprise emissions from energy production which is sourced externally. Scope 3 comprise all other emissions along the value chain. Specifically, this includes the greenhouse gas emissions of raw materials obtained from suppliers. Product Carbon Footprint (PCF) sum up greenhouse gas emissions and removals from the consecutive and interlinked process steps related to a particular product. Cradle-to-gate PCF sum up greenhouse gas emissions based on selected process steps: from the extraction of resources up to the factory gate where the product leaves the company. Such PCFs are called partial PCFs. In order to achieve such summation, each company providing any products must be able to provide the scope 1 and scope 2 contributions to the PCF for each of its products as accurately as possible and obtain reliable and consistent data for the PCFs of purchased energy (scope 2) and their raw materials (scope 3).

In an embodiment, chemical property may be a property established by changing the chemical structure, in particular of a material. The chemical product may be obtained by changing the chemical structure of one or more educts. Chemical property of the chemical product may include properties associated with a chemical reaction of the chemical product. Examples may include reactivity, electronegativity or the like. Physical property may be one of the following: mechanical properties, electrical properties, optical properties, thermal properties or the like. For example, physical property may comprise one or more of the following density, scratch resistance, electrical

conductivity, color, absorption, heat capacity or the like. Biological property may include a property related to an activity of a living organism.

In an embodiment, task instruction may include at least a part of the request, at least a part of the functional specification data and/or at least a part of the one or more input data structure(s). The task instruction may be generated by combining at least the part of the request, at least the part of the functional specification data and/or at least the part of the one or more input data structure(s). The task instruction may be provided to the one or more data-driven model(s). The one or more data-driven model(s) may be configured to select at least one operating engine from the one or more operating engine(s). The selected operating engine may be associated with one or more function(s) for providing the chemical product with the one or more target properties, in particular for providing a digital representation of the chemical structure of the chemical product with the one or more target properties. In an embodiment, the task instruction may comprise a selection task instruction and a structure task instruction. The selection task instruction may be generated by combining at least a part of the request and the functional specification data. The selection task instruction may be provided to a selection model for selecting at least one operating engine. The selection model may provide model output data in response to being provided with the selection task instruction. The model output data may be associated and/or indicative of the selected operating engine and at least a part of the request, in particular, the one or more target properties. The structure task instruction may be generated by merging at least a part of the model output data and the one or more input data structure(s). The structure task instruction may be provided to a structure model for generating operating input data, in particular structured operating input data. The structure model may provide operating input data in response to being provided with the structure task instruction.

In an embodiment, providing the task instruction may include mapping the task instruction to numerical representation of the task instruction. A numerical representation of the task instruction may be the vectorized task instruction. Hence, the terms vectorized task instruction and numerical representation of the task instruction may be used interchangeably. The numerical representation of the task instruction may comprise a tensor such as a matrix and/or a vector. The one or more data-driven model(s) may be configured to map the numerical representation of the task instruction to numerical representation of the operating input data, and to map the numerical representation of the operating input data to operating input data. The numerical representation of the task instruction may be related to and/or may represent at least a part of the request, at least a part of the functional specification data and/or at least a part of the one or more input data structure(s). The numerical representation of the task instruction may be obtained by passing the task instruction through one or more embedding layers. The one or more data-driven model(s) may include one or more embedding layers. The one or more embedding layers may be configured to map unstructured data to a structured numerical representation, in particular a numerical representation of the data. The numerical representation of the task instruction may be

indicative and/or may depend on the sequence of one or more elements and/or string data related to the task instruction. The sequence may be encoded by positional encoding of the numerical representation of the task instruction. Hence, the numerical representation of the task instruction may be mapped to a numerical representation of the task instructions and a relation between two or more parts of the task instruction. The numerical representation of the task instruction and the relation between the two or more parts of the task instruction may be mapped to the context task instruction. Positional encoding may be performed prior to providing the task instruction to the one or more data-driven model(s) or by processing of the one or more data-driven model(s), in particular the one or more embedding layer(s). Alternatively, the self-attention mechanism of the encoder may include relative positional encoding as described 1803.02155.pdf (arxiv.org). The numerical representation of the task instruction may be associated with, in particular comprise, structured numerical data, in particular a tensor, related to the request, the functional specification data and the one or more input data structure(s). In particular, the numerical representation of the task instruction may represent the task instruction, preferably at least a part of the request, the one or more input data structure(s) and/or the functional specification data. The numerical representation of the task instruction may be associated with a smaller amount of data than the task instruction. Further, the numerical representation of the task instruction may be processed by one or more matrix operation(s) associated with the one or more data-driven model(s). Hence, the numerical representation of the task instruction may be faster processable for matrix operations of the one or more data-driven model(s). The numerical representation of the task instruction may be a structured digital representation of the task instruction including unstructured data, in particular associated with a machine processable numerical, preferably float, format. The structured numerical representation of the task instruction can be efficiently processed by one or more data-driven model(s) and allows to save significant computational resources for processing unstructured requests.

The one or more data-driven model(s) may include one or more encoder block(s) and/or one or more decoder block(s) for mapping the numerical representation of the task instruction to the context task instruction. Context task instruction may be related to numerical representation of the task instruction. Context task instruction may be numerical representation of the task instruction processed by one or more matrix operation(s) associated with the one or more data-driven model(s). The one or more encoder block(s) and/or one or more decoder block(s) may be configured to map the numerical representation of the task instruction to the context task instruction, in particular by taking a sequence of one or more elements and/or string data related to the task instruction into account. The context task instruction may be associated with, in particular comprise, structured numerical data, in particular a tensor. Context task instruction may represent a sequence of elements associated with the task instruction and a relation between the elements of the sequence. The relation between the elements may be obtained by applying the one or more matrix operation(s) to the numerical representation of the task instruction. The elements may comprise at least a part of a word, a number, a symbol or the like. Thereby, the relation

between the elements, e.g. words in a text, can be understood by the one or more data-driven model(s). This improves the mapping between the task instruction and the operating input data. Consequently, the operating engine(s) can be operated more efficiently to process the request.

5 Further, the one or more data-driven model(s) may comprise one or more encoder output(s), in particular where the one or more data-driven model(s) may include one or more encoder block(s). Further, the one or more data-driven model(s) may comprise one or more decoder output(s), in particular where the one or more data-driven model(s) may include one or more decoder block(s). Additionally or alternatively, the one or more encoder output(s) and/or the one or more decoder output(s) may be configured to map the context task instruction to a
10 plurality of confidence scores associated with a plurality of elements, in particular a distribution of confidence scores associated with the plurality of elements. At least a part of the plurality of the element(s) may be associated with, in particular included by, the operating input data. The one or more encoder block(s) and/or decoder block(s) may generate a distribution of confidence scores associated with the plurality of elements comprising one or more element(s) associated with the operating input data. The operating input data may be
15 determined by selecting the one or more element(s) associated with the operating input data according to the one or more confidence score(s) associated with the one or more element(s). In an embodiment, the one or more encoder output(s) and/or one or more decoder output(s) may be configured to select the one or more element(s) associated with the operating input data according to the one or more confidence score(s) associated with the one or more element(s). Selecting the one or more element(s) associated with the operating input data according to
20 the one or more confidence score(s) associated with the one or more element(s) may comprise receiving a range of confidence scores and selecting the one or more element(s) associated with one or more confidence score(s) within the range. The one or more encoder block(s) and/or one or more encoder output(s) and/or one or more decoder output(s) may map the vectorized task instruction to operating input data, preferably to a distribution of confidence scores associated with the plurality of elements comprising one or more elements associated with the
25 operating input data. The one or more decoder block(s) and one or more decoder output(s) may map the vectorized task instruction to operating input data, preferably to a distribution of confidence scores associated with the plurality of elements comprising one or more elements associated with the operating input data.

In an embodiment, the unstructured data, in particular the unstructured request, the unstructured task instruction,
30 the unstructured functional specification data, the unstructured model output data and/or the unstructured chemical product data, may include string data and/or a sequence of one or more elements. An element may comprise a number, a letter, a symbol or the like. Humans need to understand and intervene into decisions of machines, in particular in fields with high safety requirements and humans as domain experts. This disclosure allows for translating machine interpretable data into human interpretable data. Further, an intervention and/or
35 triggering with human interpretable data input is enabled. Hence, this disclosure enables trustworthy AI.

In an embodiment, generating operating input data for one or more selected operating engine(s) in relation to the provided one or more target properties may include selecting at least one operating engine from the one or more operating engine(s) based on the task instruction by the one or more data-driven model(s), in particular the request and the functional specification data, and/or structuring the task instruction for generating operating input data by the one or more data-driven model(s). At least one of the one or more data-driven model(s) may be configured to select at least one operating engine from the one or more operating engine(s) based on the task instruction, in particular the request and the functional specification data, and structure the task instruction for generating operating input data. Additionally or alternatively, the one or more data-driven model(s) may comprise two or more data-driven model(s). The two or more data-driven model(s) may be configured to perform at least one of select at least one operating engine from the one or more operating engine(s) based on the task instruction, in particular the request and the functional specification data, or structure the task instruction for generating operating input data per data-driven model. Task-specific models are more accurate than general models as they can be tailored to fulfill the one or more predefined function(s). The two prone processing of the request allows to reliably generate operating input data. This in turn improves the robustness of generating the chemical product data and hence, the robustness of providing the chemical product with the one or more target properties.

In an embodiment, the one or more data-driven model(s) may be task specific, partially task agnostic or task agnostic. The selection model, the validation model, the structure model and/or the processing model may be data-driven model(s). In an embodiment, the selection model, the validation model, the structure model and/or the processing model may be task specific, partially task agnostic or task agnostic. The one or more task specific model(s) may be configured to perform one of

- select at least one operating engine from the one or more operating engine(s) or
- structure the task instruction to generate operating input data or
- classify if a data structure related to the operating input data corresponds to the input data structure related to the selected operating engine or
- generate unstructured data in response to being provided with at least partially structured processing task instruction

per model. The one or more partially agnostic data-driven model(s) may be configured to perform one or more of

- select at least one operating engine from the one or more operating engine(s) or
- structure the task instruction to generate operating input data or
- classify if the data structure related to the operating input data corresponds to the input data structure related to the selected operating engine or

generate unstructured data in response to being provided with at least partially structured processing task instruction

per model. In particular, at least one of the one or more partially agnostic model(s) may be configured to perform at least two of

5 select at least one operating engine from the one or more operating engine(s) or

structure the task instruction to generate operating input data or

classify if the data structure related to the operating input data corresponds to the input data structure related to the selected operating engine or

10 generate unstructured data in response to being provided with at least partially structured processing task instruction

per model. In an embodiment, the one or more agnostic model(s) may include at least one data-driven model configured to

select at least one operating engine from the one or more operating engine(s) and

structure the task instruction to generate operating input data and

15 classify if the data structure related to the operating input data corresponds to the input data structure related to the selected operating engine and

generate unstructured data in response to being provided with at least partially structured processing task instruction. In an embodiment, the one or more data-driven model(s) may include a structure data-driven model, a selection model, a validation model and/or a processing model. At least partially agnostic models

20 provide the advantage of requiring less models for generating the chemical product data. Thus, computational resources for building and maintaining a plurality of models can be saved or used to obtain more accurate at least partially agnostic models. Using task-specific models in turn allows for more accurate and robust generation of operating input data. This in turn improves the robustness of generating the chemical product data and hence, the robustness of providing the chemical product with the one or
25 more target properties.

In an embodiment, providing the task instruction to the one or more data-driven model(s) may comprise

providing a selection task including unstructured data related to the request and the functional specification data to a selection model configured to generate model output data related to the selected operating engine

30 and the one or more target properties, wherein model output data includes unstructured data, and,

providing the generated model output data to a structure model configured to generate operating input data from the model output data.

In particular, providing the task instruction to the one or more data-driven model(s) may comprise

providing selection task instruction related to the request and the functional specification data to a selection model for selecting at least one operating engine from the one or more operating engine(s), wherein the selection task instruction may include unstructured data, and wherein the selection model may be configured to generate model output data related to the selected operating engine and the one or more target properties, and

providing structure task instruction related to the model output data and the one or more input data structure(s) to a structure model for generating the operating input data, wherein the structure model may be configured to generate operating input data associated with the selected operating engine and the one or more target properties from the structure task instruction. The structure task instruction may be associated with instructions that configure the structure model to generate structure operating input data. The selection task instruction may be associated with instructions that configure the selection model to generate structure model output data. The selection task instruction may include at least a part of the request, in particular the one or more target properties. Further, the selection task instruction may be derived from the request and/or the request may depend on the one or more target properties. The selection task instruction may further include at least a part of the functional specification data. The model output data may be related to the at least one selected operating engine and/or the one or more target properties. Preferably, the model output data may be indicative of the at least one selected operating engine. The model output data may include the one or more target properties. Separating a task into two tasks allows to use task-specific data-driven models with one or more predefined function(s). Usually, task-specific models are more accurate than general models as they can be tailored to fulfill the one or more predefined function(s). The two prone processing of the request allows to reliably generate operating input data. This in turn improves the robustness of generating the chemical product data and hence, the robustness of providing the chemical product with the one or more target properties.

In an embodiment, providing the selection task instruction to the selection model may include mapping the selection task instruction to a numerical representation of the selection task instruction. The selection model may be configured to map the numerical representation of the selection task instruction to a numerical representation of the model output data, and to map the numerical representation of the model output data to model output data.

The selection model may include one or more encoder block(s) and/or one or more decoder block(s) for mapping the numerical representation of the selection task instruction to the numerical representation of the model output data. The one or more encoder block(s) and/or one or more decoder block(s) may be configured to map the numerical representation of the selection task instruction to the numerical representation of the model output data, in particular by taking a sequence of one or more elements and/or string data related to the selection task instruction into account. Further, the selection model may comprise one or more encoder output(s), in particular

where the selection model may include one or more encoder block(s). Further, the selection model may comprise one or more decoder output(s), in particular where the selection model may include one or more decoder block(s).

5 The selection model may include one or more encoder block(s) and/or one or more decoder block(s) for mapping the numerical representation of the selection task instruction to the context selection task instruction. Context selection task instruction may be related to numerical representation of the selection task instruction. Context selection task instruction may be numerical representation of the selection task instruction processed by one or more matrix operation(s) associated with the selection model. The selection model may be further configured to map a numerical representation of the selection task instruction to a context selection task instruction and map the
10 context selection task instruction to a numerical representation of the operating input data. The context selection task instruction may be obtained by processing the numerical representation of the selection task instruction by one or more matrix operation(s) associated with the selection model. The one or more encoder block(s) and/or one or more decoder block(s) may be configured to map the numerical representation of the selection task instruction to the context selection task instruction, in particular by taking a sequence of one or more elements
15 and/or string data related to the selection task instruction into account. The context selection task instruction may be associated with, in particular comprise, structured numerical data, in particular a tensor. Context selection task instruction may represent a sequence of elements associated with the selection task instruction and a relation between the elements of the sequence. The relation between the elements may be obtained by applying the one or more matrix operation(s) to the numerical representation of the selection task instruction. The elements may
20 comprise at least a part of a word, a number, a symbol or the like. Thereby, the relation between the elements, e.g. words in a text, can be understood by the selection model. This improves the mapping between the selection task instruction and the model output data. Consequently, the operating engine(s) can be operated more efficiently to process the request.

25 Further, the selection model may comprise one or more encoder output(s), in particular where the selection model may include one or more encoder block(s). Further, the selection model may comprise one or more decoder output(s), in particular where the selection model may include one or more decoder block(s). Additionally or alternatively, the one or more encoder output(s) and/or the one or more decoder output(s) may be configured to map the context selection task instruction to a plurality of confidence scores associated with a plurality of
30 elements, in particular a distribution of confidence scores associated with the plurality of elements. At least a part of the plurality of the element(s) may be associated with, in particular included by, the model output data. The one or more encoder block(s) and/or decoder block(s) may generate a distribution of confidence scores associated with the plurality of elements comprising one or more element(s) associated with the model output data. The model output data may be determined by selecting the one or more element(s) associated with the model output
35 data according to the one or more confidence score(s) associated with the one or more element(s). In an

embodiment, the one or more encoder output(s) and/or one or more decoder output(s) may be configured to select the one or more element(s) associated with the model output data according to the one or more confidence score(s) associated with the one or more element(s). Selecting the one or more element(s) associated with the model output data according to the one or more confidence score(s) associated with the one or more element(s) may comprise receiving a range of confidence scores and selecting the one or more element(s) associated with one or more confidence score(s) within the range. The one or more encoder block(s) and/or one or more encoder output(s) and/or one or more decoder output(s) may map the numerical representation of the selection task instruction to model output data, preferably to a distribution of confidence scores associated with the plurality of elements comprising one or more elements associated with the model output data. The one or more decoder block(s) and one or more decoder output(s) may map the numerical representation of the selection task instruction to model output data, preferably to a distribution of confidence scores associated with the plurality of elements comprising one or more elements associated with the operating input data.

The numerical representation of the model output data may represent the model output data, in particular the sequence of the plurality of elements and/or the string data related to the model output data. The numerical representation of the model output data may comprise a numerical representation of the model output data. The numerical representation of the selection task instruction may be indicative and/or may depend on the sequence of one or more elements and/or string data related to the selection task instruction. The numerical representation of the selection task instruction may be related to and/or may represent at least a part of the request and/or at least a part of the functional specification data. The numerical representation of the selection task instruction may be obtained by passing the selection task instruction through one or more embedding layers. Hence, the one or more selection model may include one or more embedding layers. The one or more embedding layers may be configured to map unstructured data to a structured numerical representation, in particular vectorized data. The numerical representation of the selection task instruction may be indicative and/or may depend on the sequence of one or more elements and/or string data related to the selection task instruction. The sequence may be encoded by positional encoding of the numerical representation of the selection task instruction. Positional encoding may be performed prior to providing the selection task instruction to the structure model or by processing of the structure model, in particular the one or more embedding layer(s). Alternatively, the self-attention mechanism of the encoder may include relative positional encoding as described 1803.02155.pdf (arxiv.org). The numerical representation of the selection task instruction may comprise structured numerical data, in particular a tensor, related to the request and the functional specification data. In particular, the numerical representation of the selection task instruction may represent the selection task instruction, preferably at least a part of the request and the functional specification data. The numerical representation of the selection task instruction may be associated with a smaller amount of data than the selection task instruction. Further, the numerical representation of the selection task instruction may be processed faster by the selection model, in

particular by one or more matrix operation(s). The numerical representation of the selection task instruction may require less computational storage than the selection task instruction. The numerical representation of the selection task instruction may be a structured digital representation of the selection task instruction including unstructured data. The numerical representation of the selection task instruction can be efficiently processed by the selection model and allows to save significant computational resources for processing unstructured requests. Thereby, the chemical product data can be generated reliably by the selected operating engine.

In an embodiment, providing the structure task instruction to the structure model may include mapping the structure task instruction to a numerical representation of the structure task instruction. The structure model may be configured to map the numerical representation of the structure task instruction to a numerical representation of the operating input data, and to map the numerical representation of the operating input data to operating input data. The numerical representation of the operating input data may be associated with a distribution of confidence scores associated with the plurality of elements comprising one or more elements associated with the operating input data.

The numerical representation of the structure task instruction may be obtained by passing the structure task instruction through one or more embedding layers. The structure model may include one or more embedding layers. The one or more embedding layers may be configured to map unstructured data to a structured numerical representation, in particular vectorized data. The numerical representation of the structure task instruction may be indicative and/or may depend on the sequence of one or more elements and/or string data related to the structure task instruction. The sequence may be encoded by positional encoding of the numerical representation of the structure task instruction. Positional encoding may be performed prior to providing the structure task instruction to structure model or by processing of the structure model, in particular the one or more embedding layer(s). Alternatively, the self-attention mechanism of the encoder may include relative positional encoding as described 1803.02155.pdf (arxiv.org). In particular, the numerical representation of the task instruction may represent the structure task instruction. The numerical representation of the structure task instruction may be associated with a smaller amount of data than the structure task instruction. Further, the numerical representation of the structure task instruction may be processed by one or more matrix operation(s) associated with the structure model. Hence, the numerical representation of the structure task instruction may be faster processable for matrix operations of the structure model. The numerical representation of the structure task instruction may be a structured digital representation of the structure task instruction including unstructured data, in particular associated with a machine processable numerical, preferably float, format. The structured numerical representation of the task instruction can be efficiently processed by the structure model and allows to save significant computational resources for processing unstructured requests.

In an embodiment, the structure model may include one or more encoder block(s) and/or one or more decoder block(s) for mapping the numerical representation of the structure task instruction to the context structure task instruction. Context structure task instruction may be related to a numerical representation of the structure task instruction. Context structure task instruction may be a numerical representation of the structure task instruction obtained by processing the numerical representation of the structure task instruction by one or more matrix operation(s) associated with the structure model. The one or more encoder block(s) and/or one or more decoder block(s) may be configured to map the numerical representation of the structure task instruction to the context structure task instruction, in particular by taking a sequence of one or more elements and/or string data related to the task instruction into account. The context structure task instruction may be associated with, in particular comprise, structured numerical data, in particular a tensor. Context structure task instruction may represent a sequence of elements associated with the structure task instruction and a relation between the elements of the sequence. The relation between the elements may be obtained by applying the one or more matrix operation(s) to the numerical representation of the structure task instruction. The elements may comprise at least a part of a word, a number, a symbol or the like. Thereby, the relation between the elements, eg words in a text, can be understood by the structure model. This improves the mapping between the structure task instruction and the operating input data. Consequently, the operating engine(s) can be operated more efficiently to process the request.

Further, the structure model may comprise one or more encoder output(s), in particular where the structure model may include one or more encoder block(s). Further, the structure model may comprise one or more decoder output(s), in particular where the structure model may include one or more decoder block(s). Additionally or alternatively, the one or more encoder output(s) and/or the one or more decoder output(s) may be configured to map the context structure task instruction to a plurality of confidence scores associated with a plurality of elements, in particular a distribution of confidence scores associated with the plurality of elements. At least a part of the plurality of the element(s) may be associated with, in particular included by, the operating input data. The one or more encoder block(s) and/or decoder block(s) may generate a distribution of confidence scores associated with the plurality of elements comprising one or more element(s) associated with the operating input data. The operating input data may be determined by selecting the one or more element(s) associated with the operating input data according to the one or more confidence score(s) associated with the one or more element(s). In an embodiment, the one or more encoder output(s) and/or one or more decoder output(s) may be configured to select the one or more element(s) associated with the operating input data according to the one or more confidence score(s) associated with the one or more element(s). Selecting the one or more element(s) associated with the operating input data according to the one or more confidence score(s) associated with the one or more element(s) may comprise receiving a range of confidence scores and selecting the one or more element(s) associated with one or more confidence score(s) within the range. The one or more encoder block(s)

and/or one or more encoder output(s) and/or one or more decoder output(s) may map the numerical representation of the structure task instruction to operating input data, preferably to a distribution of confidence scores associated with the plurality of elements comprising one or more elements associated with the operating input data. The one or more decoder block(s) and one or more decoder output(s) may map the numerical representation of the structure task instruction to operating input data, preferably to a distribution of confidence scores associated with the plurality of elements comprising one or more elements associated with the operating input data.

In an embodiment, any one of the methods may further comprise generating the selection task instruction by combining the request and the functional specification data and/or generating the structure task instruction by merging the one or more input data structure(s) and the model output data. The structure task instruction may be further associated with instructions for triggering the structure model and/or the one or more data-driven model(s) to generate the operating input data. By combining provided and/or generated data, the available context can be provided to the selection model and/or the structure model to accurately select the at least one operating engine and/or structure the operating input data. Thereby, the chemical product data can be reliably generated.

In an embodiment, any one of the methods may further comprise providing a validation task instruction related to the request and the operating input data to a validation model for validating the at least one selected operating engine corresponds to the request. The validation model may be configured to follow task instructions. Optionally, the validation model may be the selection model.

In an embodiment, any one of the methods may further comprise providing a validation task instruction related to the operating input data, the at least one selected operating engine and the one or more input data structure(s), in particular the at least one input data structure(s) associated with the at least one selected operating engine, to a validation model for validating a data structure related to the operating input data. The validation model may be configured to classify if the data structure related to the operating input data may correspond to the input data structure related to the at least one selected operating engine. The validation model may be configured to provide an indication on whether the data structure related to the operating input data corresponds to the input data structure related to the at least one selected operating engine. The indication may be a class label indicating whether the data structure related to the operating input data may be validated. Hence, the validation model may be a classification model. The classification model may be trained to provide the indication in response to receiving the validation task instruction. The validation model may comprise one or more classification layers. The one or more classification layers may be configured to determine the indication on whether the data structure related to the operating input data corresponds to the input data structure related to the at least one selected operating engine from the validation task instruction, in particular the numerical representation of the validation

task instruction. Hence, the indication on whether the data structure related to the operating input data corresponds to the input data structure related to the at least one selected operating engine may be derived from and/or may depend on the validation task instruction. Validating the data structure related to the operating input data ensures robust triggering of the at least one selected operating engine by the operating input data. Thereby, the chemical product data can be reliably generated. Using a task specific model for validating the data structure related to the operating input data allows for a high accuracy of the validating.

In an embodiment, providing the validation task instruction may comprise generating validation task instruction by merging the one or more input data structure(s), at least a part of the model output data and the operating input data and providing the validation task instruction data to the validation model and/or the one or more data-driven model(s). The validation task instruction may comprise the one or more input data structure(s), an indication on the selected operating engine and the operating input data. Further, the validation task instruction may include instructions for triggering the validation model and/or the one or more data-driven model(s) to validate the data structure associated with the operating input data. By combining provided and/or generated data, the available context can be provided to the validation model and/or the one or more data-driven model(s). This allows to accurately validate the operating input data. This ensures robust triggering of the at least one selected operating engine by the operating input data. Thereby, the chemical product data can be reliably generated.

In an embodiment, providing the validation task instruction may include mapping the validation task instruction to a numerical representation of the validation task instruction. The validation model may be configured to map the numerical representation of the validation task instruction to a numerical representation of the indication on whether the data structure related to the operating input data corresponds to the input data structure related to the at least one selected operating engine. The validation model may be further configured to map the numerical representation of the indication to the indication on whether the data structure related to the operating input data corresponds to the input data structure related to the at least one selected operating engine.

The numerical representation of the validation task instruction may be obtained by passing the validation task instruction through one or more embedding layers. The validation model may include one or more embedding layers. The one or more embedding layers may be configured to map unstructured data to a structured numerical representation, in particular vectorized data. The numerical representation of the validation task instruction may be indicative and/or may depend on the sequence of one or more elements and/or string data related to the validation task instruction. The sequence may be encoded by positional encoding of the numerical representation of the validation task instruction. Positional encoding may be performed prior to providing the validation task instruction to the validation model or by processing of the validation model, in particular the one or more embedding layer(s). Alternatively, the self-attention mechanism of the encoder may include relative positional

encoding as described 1803.02155.pdf (arxiv.org). The numerical representation of the validation task instruction may be associated with, in particular comprise, structured numerical data. In particular, the numerical representation of the validation task instruction may represent the validation task instruction. The numerical representation of the validation task instruction may be associated with a smaller amount of data than the validation task instruction. Further, the numerical representation of the validation task instruction may be processed by one or more matrix operation(s) associated with the validation model. Hence, the numerical representation of the validation task instruction may be faster processable for matrix operations of the validation model. The numerical representation of the validation task instruction may be a structured digital representation of the validation task instruction including unstructured data, in particular associated with a machine processable numerical, preferably float, format. The structured numerical representation of the validation task instruction can be efficiently processed by the validation model and allows to save significant computational resources for processing unstructured requests.

The validation model may include one or more encoder block(s) and/or one or more decoder block(s) for mapping the numerical representation of the validation task instruction to the context validation task instruction. Context validation task instruction may be related to a numerical representation of the validation task instruction. Context validation task instruction may be obtained by processing the numerical representation of the validation task instruction by one or more matrix operation(s) associated with the validation model. The one or more encoder block(s) and/or one or more decoder block(s) may be configured to map the numerical representation of the validation task instruction to the context validation task instruction, in particular by taking a sequence of one or more elements and/or string data related to the validation task instruction into account. The context validation task instruction may be associated with, in particular comprise, structured numerical data, in particular a tensor. Context validation task instruction may represent a sequence of elements associated with the validation task instruction and a relation between the elements of the sequence. The relation between the elements may be obtained by applying the one or more matrix operation(s) to the numerical representation of the validation task instruction. The elements may comprise at least a part of a word, a number, a symbol or the like. Thereby, the relation between the elements, e.g. words in a text, can be understood by the one or more data-driven model(s). This improves the mapping between the validation task instruction and the indication on whether the data structure related to the operating input data corresponds to the input data structure related to the at least one selected operating engine. Consequently, the operating engine(s) can be operated more efficiently to process the request.

Further, the validation model may comprise one or more encoder output(s), in particular where the validation model may include one or more encoder block(s). Further, the validation model may comprise one or more decoder output(s), in particular where the one or more data-driven model(s) may include one or more decoder

block(s). Additionally or alternatively, the one or more encoder output(s) and/or the one or more decoder output(s) may be configured to map the context validation task instruction to a plurality of confidence scores associated with a plurality of elements, in particular a distribution of confidence scores associated with the plurality of elements. At least a part of the plurality of the element(s) may be associated with, in particular included by, the indication. The one or more encoder block(s) and/or decoder block(s) may generate a distribution of confidence scores associated with the plurality of elements comprising one or more element(s) associated with the indication. The indication may be determined by selecting the one or more element(s) associated with the indication according to the one or more confidence score(s) associated with the one or more element(s). In an embodiment, the one or more encoder output(s) and/or one or more decoder output(s) may be configured to select the one or more element(s) associated with the indication according to the one or more confidence score(s) associated with the one or more element(s). Selecting the one or more element(s) associated with the indication according to the one or more confidence score(s) associated with the one or more element(s) may comprise receiving a range of confidence scores and selecting the one or more element(s) associated with one or more confidence score(s) within the range.

The one or more encoder block(s) and/or one or more encoder output(s) and/or one or more decoder output(s) may map the numerical representation of the validation task instruction to the indication, preferably to a distribution of confidence scores associated with the plurality of elements comprising one or more elements associated with the indication. The one or more decoder block(s) and one or more decoder output(s) may map the numerical representation of the validation task instruction to the indication, preferably to a distribution of confidence scores associated with the plurality of elements comprising one or more elements associated with the indication. The numerical representation of the validation task instruction may be indicative and/or may depend on the sequence of one or more elements and/or string data related to the validation task instruction. The indication may comprise string data indicative of whether the data structure related to the operating input data corresponds to the input data structure related to the at least one selected operating engine. The numerical representation of the indication may comprise numerical data, in particular structured numerical data. The numerical representation of the indication may represent the indication, in particular the sequence related to the indication, in particular the indication including unstructured data. The one or more encoder output(s) and/or the one or more decoder output(s) may be further configured to map the numerical representation of the indication to the indication related to a sequence of a plurality of elements or including string data.

In an embodiment, the at least one selected operating engine may comprise a database configured to provide a digital representation of a chemical structure of the chemical product with the one or more target properties in response to receiving a structured query related to the one or more target properties. The operating input data may comprise the structured query. The chemical product data may comprise the digital representation. At least a

part of the chemical product data may be retrieved by querying a structured database. Structured databases may provide data reliably and reliably. Followingly, using a structure database as the at least one operating engine allows for reliable and reliable generation of chemical product data. Thereby, the decisions arising based on this disclosure are reasonable. This enables an improved human-machine interaction for providing the chemical product with the one or more target properties.

In an embodiment, the at least one selected operating engine may comprise a database configured to be provided with the operating input data and determining a digital representation of a chemical structure of the chemical product with the one or more target properties corresponding to the operating input data by

mapping the operating input data to a numerical representation of the operating input data

determining one or more distance(s) between the numerical representation of the operating input data and a numerical representation of the digital representations of the chemical structure of two or more chemical products, wherein the numerical representation of the digital representations are obtained by mapping the digital representations of the chemical structure of the two or more chemical products to the numerical representation of the digital representations of the chemical products

selecting the digital representation of the chemical structure of the chemical product with the one or more target properties by determining the digital representation of the chemical structure associated with the smallest distance. The chemical product data may comprise the digital representation. The embedding database may allow to retrieve at least a part of the chemical product data independent of keywords and thus, more accurately according to the context of the request. Hence, more accurate chemical product data can be generated. The numerical representation of the digital representations of the chemical structure of the one or more chemical product(s) may be obtained by using one or more embedding layer(s) according to SMILESVec or Mol2Vec.

In an embodiment, the at least one selected operating engine may include one or more subengine(s) configured to

select at least one subengine from the one or more subengine(s) based on the operating input data and subengine specification data related to one or more functions of the one or more subengine(s) for providing the chemical product with the one or more target properties,

and structure subengine task instruction related to the operating input data and one or more input data structure(s) related to the one or more subengine(s) to generate subengine input data to the at least one selected subengine,

and optionally classify if the data structure related to the subengine input data corresponds to the input data structure related to the at least one selected subengine

and provide at least a part of the chemical product data, preferably the operating output data, in response to providing the subengine input data.

The subengine input data may be derived from the operating input data and/or may depend on the operating input data. The subengine input data may include at least a part of the operating input data. The at least one subengine may provide at least a part of the chemical product data, preferably the operating output data, in response to providing the subengine input data. In an embodiment, the one or more subengine(s) may perform at least one of

5 select at least one subengine from the one or more subengine(s) based on the operating input data and subengine specification data related to one or more functions of the one or more subengine(s), structure subengine task instruction related to the operating input data and one or more input data structure(s) related to the one or more subengine(s) to generate subengine input data to the at least one

10 selected subengine, optionally classify if the data structure related to the subengine input data corresponds to the input data structure related to the at least one selected subengine, provide at least a part of the chemical product data, preferably the operating output data, in response to providing the subengine input data

15 or a combination thereof per subengine. This provides the advantage to separate tasks into a plurality of tasks. By doing so, intermediate steps become obvious. Hence, this feature allows for reasoning of the decision taken according to this disclosure. This in turn reduced the errors of processing the received request. Hence, this contributes to a robust generation of the chemical product data and enables trustworthy AI.

20 In an embodiment, generating chemical product data in response to providing the operating input data may comprise

providing a digital representation of a chemical structure of one or more educts, determining a digital representation of a chemical structure of one or more chemical products formed in one

25 or more chemical reaction(s) of the one or more educts, determining one or more properties associated with the one or more chemical products by providing the digital representation of the chemical structure of the one or more chemical products to a property model, wherein the property model is configured to be provided with digital representations of chemical products and providing one or more properties associated with the chemical products,

30 selecting the chemical product with the one or more target properties by comparing the properties associated with the one or more chemical products, and providing the digital representation of the chemical structure of the chemical product associated with the target property. The chemical product data may comprise and/or represent the digital representation. Further including determining one or more formation score(s) associated with forming the one or more

35 chemical products in the one or more chemical reaction(s) of the one or more educt(s). The formation score

may be determined by providing the digital representation of the chemical structure of the one or more educt(s) and/or of the one or more chemical product(s) to a scoring data-driven model. The scoring data-driven model may be configured to provide the one or more formation score(s) in response to providing the digital representation of the chemical structure of the one or more educt(s) and/or of the one or more chemical product(s). The chemical product with the one or more target properties may be further selected based on the one or more formation score(s). The one or more formation score(s) may be compared with a predefined range related to forming the chemical product with the one or more target properties. The formation score allows to judge on the efficiency of producing the one or more chemical products by the one or more chemical reaction(s). Hence, taking the formation score into account for selecting the chemical products with the one or more target properties allows for selecting chemical products that can be synthesized with a low resource investment.

In an embodiment, the chemical product data may further comprise unstructured data. Generating the chemical product data may further comprise providing processing task instruction related the output of the processing model, in particular the digital representation of the chemical structure of the one or more product(s), preferably chemical product(s), and the request to a processing model. The processing model may be configured to generate unstructured data in response to being provided with data including at least partially structured data. The processing task instruction may be obtained by merging the digital representation and the request. The processing task instruction may include instructions for triggering the processing model to generate the chemical product data from the operating output data and/or the digital representation of the chemical structure of the one or more chemical product(s) and the request. Further, the processing task instruction may include at least a part of the operating output data, in particular the digital representation of the chemical structure of the one or more product(s) and the request. The chemical product data may be provided via a user interface. Additionally or alternatively, the processing task instruction may be provided to the one or more data-driven model(s). The one or more data-driven model(s) may be further configured to generate unstructured data in response to being provided with data including at least partially structured data. Providing chemical product data including unstructured data allows to present the chemical product data in a human-interpretable format. Hence, the human-machine interaction is improved by allowing for verifying the generated chemical product data with domain expert knowledge and trustworthy AI is enabled.

In an embodiment, providing the processing task instruction may comprise merging the request and the model output data and/or the digital representation of the chemical structure of the chemical product with the one or more target properties and providing the processing task instruction to the processing model. Combining may refer to merging the request and the model output data and/or the digital representation of the chemical structure of the chemical product with the one or more target properties. Combining the request and the model output data

and/or the digital representation of the chemical structure may result in processing task instruction including the request and the model output data and/or the digital representation of the chemical structure. By combining provided and/or generated data, the available context can be provided to the processing model and/or the one or more data-driven model(s). Thereby, the chemical product data can be reliably generated.

5

In an embodiment, the processing task instruction may be related to at least the part of the chemical product data and the request. At least the part of the chemical product data may be the operating output data, i.e. as provided by the at least one operating engine. The chemical product data may comprise the operating output data and/or may be indicative of the operating output data. The chemical product data may correspond to the request if the at least the part of the chemical product data, in particular the operating output data, is associated with providing the target chemical product. In particular, the chemical product data may correspond to the request if providing at least the part of the chemical product data, in particular the operating output data, to a control and/or monitoring engine of a chemical production facility may trigger production and/or a monitoring of a processing and/or a production of the target chemical product. In particular, the chemical product data may correspond to the request if at least the part of the chemical product data, in particular the operating output data, may be configured to trigger a control engine of a chemical production facility to produce the target chemical product and/or a monitoring engine of the chemical production facility to monitor a processing and/or the production of the target chemical product.

10

15

In an embodiment, the processing model may be configured to provide and/or generating the chemical product data in response to determining that at least the part of the chemical product data, in particular the operating output data, may correspond to the request. Otherwise, the processing model may be configured to provide processing data related to the request, in particular indicative of at least a part of the request independent of at least the part of the chemical product data. The processing data, the request and the functional specification data may be provided to the selection model for evaluating the selection of the at least one operating engine. The selection model may select at least one operating engine different from the previously selected operating engine. Operating input data associated with the at least one operating engine different from the previously selected operating engine may be generated and/or provided to the at least one operating engine. By introducing a correction loop, potential errors can be directly identified and remedied. Hence, the accuracy of provided chemical product data is increased. Ultimately, this contributes to improving the efficiency of monitoring and/or controlling production and/or processing of chemical products.

20

25

30

In an embodiment, providing the processing task instruction may include mapping the processing task instruction to a numerical representation of the processing task instruction. The processing model and/or the one or more data-driven model(s) may be configured to map the numerical representation of the processing task instruction to

a numerical representation of the chemical product data, and/or to map the numerical representation of the chemical product data to chemical product data. The numerical representation of the processing task instruction may be associated with a smaller amount of data than the processing task instruction. Further, the numerical representation of the processing task instruction may be processed faster by the processing model, in particular
5 by one or more matrix operation(s) related to the processing model. The numerical representation of the processing task instruction may require less computational storage than the processing task instruction. The numerical representation of the processing task instruction may be a structured digital representation of the processing task instruction including unstructured data. The numerical representation of the processing task instruction can be efficiently processed by the processing model and allows to save significant computational
10 resources for processing unstructured requests. Thereby, the chemical product data can be generated reliably. The numerical representation of the chemical product data may be associated with, preferably comprise, a distribution of confidence scores associated with the plurality of elements comprising one or more elements associated with the operating input data. The chemical product data may be determined by selecting the one or more element(s) associated with the chemical product data according to the one or more confidence score(s)
15 associated with the one or more element(s).

In an embodiment, any one of the methods may further include mapping a numerical representation of the operating input data to operating input data, e.g. by using a predefined relation between the numerical representation of the operating input data and the operating input data. The predefined relation may relate to a
20 vocabulary specifying a relation between the numerical representation of the operating input data and the operating input data.

In an embodiment, the functional specification data may include a digital representation of the one or more functions per operating engine and a relation of the one or more functions to the two or more operating engines
25 corresponding to the one or more functions. The digital representation of the one or more functions may comprise one or more type(s) of input and output pairs associated with the two or more operating engine(s). The digital representation of the one or more functions may comprise string data related to the one or more functions of the two or more operating engines.

30 In an embodiment, the two or more operating engine(s) may include at least one operating engine configured to provide the chemical product data and at least one operating engine configured to provide data of another type than the chemical product data.

In an embodiment, any one of the methods may further include mapping a numerical representation of the
35 operating input data to operating input data, e.g. by using a predefined relation between the numerical

representation of the operating input data and the operating input data. The predefined relation may relate to a vocabulary specifying a relation between the numerical representation of the operating input data and the operating input data. In an embodiment, any one of the methods may further include mapping numerical representation of the operating input data to operating input data, e.g. by using a predefined relation between the numerical representation of the operating input data and the operating input data. The predefined relation may relate to a vocabulary specifying a relation between the numerical representation of the operating input data and the operating input data.

In an embodiment, the selection model may be a pretrained data-driven model. The pretrained data-driven model may be configured to perform a plurality of different tasks according to a plurality of different task instructions. The plurality of different tasks may include selecting the at least one operating engine related to the request, in particular the at least one operating engine configured for providing the chemical product data according to the request. The plurality of different task instructions may include the selection task instruction. The pretrained data-driven model(s) may be parametrized and/or trained based on unstructured data, in particular text data and optionally numerical data such as tabular data or image data. The pretrained data-driven model(s) may be configured to perform a plurality of task. The pretrained data-driven model(s) may be configured to perform the task according to the provided task instruction. Hence, the pretrained data-driven model may be configured to be provided with a plurality of different task instructions and/or provide a plurality of different types of output data upon receiving different task instructions.

In an embodiment, the selection model may be a finetuned data-driven model, wherein the finetuned data-driven model is obtained by further training a pretrained data-driven model based on training data comprising selection task instructions and corresponding operating input data, preferably corresponding model output data, wherein the pretrained data-driven model is configured to perform a plurality of different tasks according to a plurality of different task instructions. The finetuned data-driven model(s) may be obtained by training pretrained data-driven model(s) configured to perform a plurality of tasks according to a plurality of task instructions. The finetuned data-driven model(s) may trained additionally on a training data set comprising a plurality of task instructions of one type and corresponding output data. The finetuned data-driven model may be trained additionally to provide output data of a predefined type according to the training data set. The finetuned data-driven model may be configured to be provided with a plurality of different task instructions and/or provide a plurality of different types of output data upon receiving different types of task instructions. Further, the finetuned data-driven model may be configured to provide one type of output data upon receiving one type of task instruction with a higher accuracy than providing other types of output data upon receiving other types of task instructions.

In an embodiment, unstructured data may comprise data generated independent of a predefined data schema and/or a data format. Unstructured data may comprise text data, numerical data, tabular data or the like. An example of a data schema and/or a data format may be JSON.

- 5 In an embodiment, the input data structure(s) may be indicative of a sequence of one or more datapoint(s) associated with the request. The input data structure(s) may specify and/or define the sequence of one or more datapoint(s) associated with the request. For example, the input data structures may comprise historical operating input data associated with the one or more operating engine(s), in particular the selected operating engine. Additionally or alternatively, the input data structures may be indicative of a schema of one or more datapoint(s)
- 10 associated with the request. The operating input data may comprise the sequence of the one or more datapoint(s) associated with the request.

In an embodiment, generating one or more numeric representation(s) of the data includes providing data separated by data type to one or more embedding model(s). The embedding model may be configured to map the

15 data per data type to the one or more numeric representation(s). The embedding model per data type may be configured to generate numerical data from non-numerical data by mapping non-numerical data into a multidimensional vector space. The embedding model per data type may be configured to vectorize non-numerical data, such as text data. The embedding model per data type may be configured to vectorize non-numerical data, such as text and/or image data. The embedding model may be configured to map one or more data type(s) to one

20 or more numerical representations. The embedding model may be configured to generate a joint or a shared representations of one or more data type(s), such as text, numerical and/or image data. The embedding model may be configured to generate one or more numeric representation(s) by including mappings between elements of the data it transforms. For text, such correlations may include semantics embedded in a trained probability distribution of the embedding model.

25

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

In the following, the present disclosure is further described with reference to the enclosed figures. The same reference numbers in the drawings and this disclosure are intended to refer to the same or like elements, components, and/or parts.

FIG. 1 illustrates an embodiment of an operating system of a chemical production facility 102.

FIG. 2 illustrates an embodiment of a method for obtaining a chemical product with a target property.

FIG. 3 illustrates an embodiment of a method for obtaining chemical product data associated with a chemical product.

FIG. 4 illustrates an embodiment of a method for obtaining a digital representation of a chemical product.

FIG. 5 illustrates an embodiment of an operating system 538.

FIG. 6A illustrates an embodiment of an operating engine and an executing service 108.

FIG. 6B illustrates an embodiment of an operating engine and an executing service 108.

FIG. 6C illustrates an embodiment of an operating engine and an executing service 108.

FIG. 6D illustrates an embodiment of an operating engine and an executing service 108.

FIG. 6E illustrates an embodiment of an operating engine and an executing service 108.

FIG. 6F illustrates an embodiment of an operating engine and an executing service 108.

FIG. 7 illustrates an embodiment of producing and/or processing a chemical product 714.

FIG. 8 illustrates an embodiment of producing and/or processing a chemical product 714.

- FIG. 9 illustrates an embodiment of the input and output data associated with the selection model, the structure model, the validation model, the one or more operating engine(s) and/or the processing model.
- 5 FIG. 10 illustrates an embodiment of the input and output data associated with the selection model.
- FIG. 11 illustrates an embodiment of the input and output data associated with the structure model.
- FIG. 12 illustrates an embodiment of the input and output data associated with the validation model.
- 10 FIG. 13 illustrates an embodiment of the input and output data associated with the processing model.
- FIG. 14 illustrates an embodiment of a user interface for receiving a chemical product with a target property.
- 15 FIG. 15 illustrates an embodiment of a user interface 1512 for receiving chemical product data.
- FIG. 16 illustrates an aspect of the subject matter in accordance with one embodiment.
- FIG. 17 illustrates embodiments of APIs for obtaining a chemical product with a target property, chemical
20 product data associated with a chemical product and/or a digital representation of a chemical product.
- FIG. 18 illustrates an embodiment of training an embedding layer.
- 25 FIG. 19A illustrates an embodiment of a transformer encoder architecture.
- FIG. 19B illustrates an embodiment of a transformer decoder architecture.
- FIG. 19C illustrates an embodiment of a transformer encoder-decoder architecture.
- 30 FIG. 20 illustrates an embodiment of training and/or deploying the transformer encoder, the transformer decoder and/or the transformer encoder-decoder.
- FIG. 21 illustrates an embodiment of input embedding.
- 35

FIG. 22 illustrates an embodiment of input embedding.

DETAILED DESCRIPTION

- 5 The following embodiments are mere examples for implementing the method, the system or application device disclosed herein and shall not be considered limiting.

The following embodiments are mere examples for implementing the method, the system or application device disclosed herein and shall not be considered limiting.

- 10 FIG. 1 illustrates an embodiment of an operating system of a chemical production facility 102 configured to provide a chemical product based on one or more request(s).

Chemical products are starting materials for a plurality of different end products. As a consequence, chemical products have to provide a variety of changing properties tailored to the intended end product. The production of chemical products starts with raw materials that are processed via one or more processing steps including for example chemical reactions conducted in reactors and purification steps. Typically, chemical products are obtained from two or more chemical reactions changing the chemical structure of the reactants and thus, changing the properties of the chemical product. For example, a liquid such as monoethylenglycol and a solid such as terephthalic acid can be converted to yield polyester. Polyester is a functional polymer with distinct properties depending on the educts and reaction conditions. This may lead to a polyester with different properties. Accordingly, the chemical reaction to produce polyester needs to be tailored to the desired properties of the polyester. The polyester can be applied in a variety of field such as clothing or packaging. Hence, the challenge constitutes in serving hundreds of customers with thousands of distinct chemical products obtained from a chemical production network with a plurality of production steps to result in distinct and tailored properties of the chemical products. The properties of chemical products highly depend on its chemical structure. Even a small change in the orientation of a subgroup of a molecule with hundreds of atoms results in a distinct chemical property. Therefore, a relation between a chemical structure of the chemical product and the properties of the chemical product is complex and challenging to control.

- 30 Followingly, it is of high importance to ensure optimal production conditions to result in chemical products with target properties to ensure robust functionality of the chemical products tailored to the respective application. To tailor the chemical product in this highly diverse environment, it is of high importance to have access to the digital representation of the chemical structure and chemical product data which indicates e.g. structure property relations, production conditions, other relations that have an impact on the properties of the chemical product.

However, generating digital representations and chemical product data for providing the chemical product with tailored properties. Multiple representations of the chemical structure, multiple chemical production lines with different equipment have to be chained in a sequence depending on the chemical product and the target chemical product properties.

5

This disclosure enables reliable and efficient production of chemical products by processing multiple requests at different stages of the chaining. The requests may in particular include

one or more requests for receiving the chemical product with the target property

one or more requests for receiving a digital representation of a chemical product

10

one or more request for receiving chemical product data

or a combination thereof. In particular, the requests may be provided independent of a predefined data structure required by operating engine(s) such as databases or data-driven model(s) requiring predefined data structures. Despite this unstructured nature, the disclosure enables obtaining the chemical product, the chemical product data and/or the digital representation of the chemical product. In particular, purpose-

15

specific operating engine(s) provide accurate tools with respect to their application purpose. Because of the high dependency of properties of chemical products on reaction conditions, ratio of educts, chemical structure or the like reliable and accurate chemical product data is generated if the operating engine matching the request is selected. Moreover, by using non-purpose specific models the chaining of digital tools generating production control data with the suitable equipment for producing the chemical product with the target properties is possible even based on unstructured requests. Such non-purpose specific models can be large language models. Their capabilities for processing requests with chemical context was shown by Bran et al. ("Augmenting large language models with chemistry tools" doi 2304.05376v5).

20

A chemical production facility 122 may comprise equipment for processing and/or producing chemical products.

25

The operating system of a chemical production facility 102 may comprise an equipment interface 120. The equipment of the chemical production facility 122 may be controlled and/or monitored via an equipment interface 120. Hence, the chemical production facility 122 may be monitored and/or controlled via the operating system of a chemical production facility 102. The equipment interface may receive chemical product data from an output interface 118. The chemical product data may be associated a chemical product to be produced and/or processed by the chemical production facility 122. Further, the chemical product data may be associated with production and/or processing conditions associated with the production and/or processing of the chemical product. The chemical product data may be obtained by receiving at least one of

30

one or more request for receiving the chemical product with the target property,

one or more requests for receiving a digital representation of a chemical product,

35

one or more requests for receiving chemical product data

or a combination thereof.

The one or more requests for receiving the chemical product with the target property may be associated with one or more target properties of the chemical product and user instructions for receiving the chemical product with the one or more target properties. Further, the one or more requests for receiving the chemical product with the target property may be associated with a target type of the chemical product and/or a target field of application associated with the chemical product. The user instructions for receiving the chemical product with one or more target properties may comprise string data. The user instructions for receiving the chemical product with the one or more target properties may be indicative a target quantity associated with the chemical product, a target quality associated with the chemical property, a target delivery associated with the chemical product or the like. An example for the request for receiving the chemical product may be described in the context of FIG. 14.

The one or more requests for receiving the digital representation of the chemical product may be associated with an indication of the chemical product. The indication on the target chemical product may be associated with one or more properties of the chemical product, one or more ingredients of the chemical products, a further digital representation of the chemical product such as trivial name or a commercial name, a type of the chemical product and/or the field of application associated with the chemical product. An example for the request for receiving the digital representation of the chemical product may be described in the context of FIG. 16.

The one or more requests for receiving the chemical product data associated with a chemical product may be associated with an indication of the chemical product. The indication on the target chemical product may be associated with one or more properties of the chemical product, one or more ingredients of the chemical products, a further digital representation of the chemical product such as trivial name or a commercial name, a type of the chemical product and/or the field of application associated with the chemical product. An example for the request for receiving the chemical product data may be described in the context of FIG. 15.

An intake interface 116 may be configured to receive at least one or the request. The intake interface 116 may provide the at least one request to a selection service 104. Further, the selection service may be configured to process operating instructions. The operating instructions may be associated with one or more operating engine(s). The one or more operating engine(s) may comprise the selected operating engine related to the request. In particular, the operating instructions may be associated with technical specification of the one or more operating engine(s). The technical specification of the one or more operating engine(s) may be associated with a data structure of input data to and/or output data from the one or more operating engine(s), a technical purpose associated with the one or more operating engine(s), a denotation associated with the one or more operating engine(s), a location of the one or more operating engine(s) or the like.

The selection service 104 may be configured to process the one or more requests and the operating instructions. The selection service 104 may generate model output data. The model output data may be associated with, in particular may be indicative of a selected operating engine. The selected operating engine may be configured to perform an operation related to the one or more received request. Hence, the model output data may be associated with at least a part of the technical specification of the selected operating engine. Further, the model output data may be associated with at least a part of the request, in particular the target property and optionally further the target type, the target field of application or the like.

For example, where a request for receiving a chemical product with a target property may be received, the selected operating engine may be configured to generate a chemical structure of the chemical product. This use case is further described within the context of FIG. 6F. Hence, the selection service 104 may select the operating engine suitable for processing at least a part of the request. The selection service 104 may select the operating engine based on the received request, in particular the user instructions and any one of the target property, the target field of application, the target type, the indication on the target chemical product or a combination thereof. The selection service 104 may select the operating engine by determining a similarity score per operating engine of the one or more operating engine(s). The selected operating engine may be associated with the similarity score higher than the similarity score associated with the other operating engine(s). The similarity score may correspond to a distance between a numerical representation of the request and numerical representation of the operating instructions. This may be described in further detail in the context of FIG. 2.

The selection service 104 may provide the model output data to a structure service 110. Further, the structure service may be configured to process data structures associated with the one or more operating engine(s) comprising at least one data structure associated with the selected operating engine. The data structures may be provided by a data structure repository 512. The structure service 110 may generate operating input data from the model output data and the data structures. The operating input data may be associated with a data structure suitable for being provided to the selected operating engine. The operating input data may be associated with, in particular may comprise, at least a part of the model output data. Hence, the structure service 110 may select the data structure corresponding to the selected operating engine and may structure the model output data to result in operating input data. Analogue to the selection service 104, the structure service may select the data structure associated with the selected operating engine according to the one or more similarity scores associated with the operating engine(s) associated with the data structures. The structure service 110 may provide operating input data in response to receiving the model output data.

The structure service may provide the operating input data to a validation service 106. The validation service 106 may be configured to validate the operating input data, in particular the data structure related to the operating input data. Hence, the validation service 106 may classify if the operating input data may be suitable for being provided to the selected operating engine. For this purpose, the validation service 106 may process the operating input data and the data structures associated with the operating engine(s). The validation service 106 may determine a confidence score associated operating input data being suitable for being provided to the selected operating engine. The confidence score may be determined analogue to the similarity score as described above. The validation service 106 may provide the operating input data in response to validating the operating input data to an executing service 108. The executing service 108 may generate operating output data from the operating input data. The operating output data may be associated with at least a part of chemical product data, in particular a digital representation of the chemical structure of the chemical product. The operating output data may be requested by the request received. The operating output data may comprise structured data. The executing service 108 may provide the operating output data to a chemical data generating service 112. The chemical data generating service 112 may generate chemical product data from the operating output data. The chemical product data may be associated with at least a part of the operating output data. Further, the chemical product data may be associated with system instructions. The system instructions may be related to the user instructions. The system instructions may be for example indicative of a delivery of the chemical product. Further, the chemical product data may comprise unstructured data such as string data. The chemical product data may be associated at least partially according to the received request. The chemical data generating service 112 may provide the chemical product data to the output interface 118. In an embodiment, the output interface may be configured to provide the chemical product data to the equipment interface 120. In another embodiment, the output interface may be configured to provide the chemical product data to an operating system of a chemical product processing facility 706. The chemical data generating service 112, the output interface 118, the selection service 104, the structure service 110, the validation service 106 and the executing service 106 may be described in more detail in the context of FIG. 5.

FIG. 2 illustrates an embodiment of a method for obtaining a chemical product with a target property.

As described within the context of FIG. 1, correlating a chemical product with properties may be non-trivial due to the nature of chemistry. Hence, finding a chemical product for an specific field of application usually requires many iterations of testing potential candidates. By doing so, many resources are consumed to arrive at tailored chemical products. Providing a request for receiving the chemical product data to a selection model and/or selection engine for receiving model output data associated with a selected operation enables receiving of unstructured and structured requests while providing accurate and robust chemical product data associated with

the chemical product. Further, the chemical product can be tailored to the target field of application associated with the chemical product.

5 A request for receiving a chemical product with a target property may be received 202. The request may comprise one or more target properties. The request may comprise further specifications associated with the chemical products such as a target type of chemical product or a target field of application associated with the chemical product. The target type of the chemical product may indicate for example whether the chemical product may be a polymer, an organic liquid, an inorganic salt or the like. The field of application may be indicative of conditions the chemical product and/or a product produced by processing the chemical product may be used under. Further, the 10 field of application may specify a use case for which the chemical product may be deployed. For example, the chemical product may be used for producing a shoe, in particular a shoe sole. The use case may be indicative of an application property such as damping characteristics. The damping characteristics may result from the damping characteristics of the chemical product used for producing the shoe. Hence, the target property may comprise one or more application properties. The request for receiving the chemical product with the target 15 property may comprise unstructured data such as string data comprising one or more text blocks and/or numerical data. The request may be provided by an entity for processing chemical products. The entity for processing the chemical product may comprise one or more chemical product processing facilities 704. The request may be provided by an output interface 716 of the operating system of a chemical product processing facility 706 as described within the context of FIG. 7 and FIG. 8. The request may be received by an intake interface 116 as 20 described within the context of FIG. 1, FIG. 5, FIG. 7 and/or FIG. 8.

Operating instructions associated with one or more operations carried out by one or more operating engine(s) may be received 206. The operating instructions may describe the one or more operations and/or the one or more operating engine(s). The operation instructions may be indicative of the data input to and the data output of the 25 one or more operating engine(s). Where a request for receiving the chemical product can be received, at least one selected operation from the one or more operations may be suitable for generating a chemical structure of the chemical product with the target property. Additionally or alternatively, the at least one selected operation may be suitable for providing control data to a chemical production facility 702 for producing the chemical product with the target property. The selected operation may be carried out by at least one selected operating engine of the one or 30 more operating engine(s). In an embodiment, the selected operation may result in providing a request for further data to the entity making the request for receiving the chemical product. The request for further data may be provided in response to determining that the operation providing the request for further data may be the selected operation.

The operating instructions may be stored in an operation instructions repository 506. The operation instructions repository 506 may comprise a database. Hence, the operation instructions may be retrieved from the operation instructions repository 506, e.g. by providing a query for receiving the operation instructions to the operation instructions repository 506. Preferably, the query for receiving the operation instructions may be provided to the operation instructions repository 506 in response to receiving the request for receiving the chemical product with the target property.

The operation instructions may comprise unstructured data associated with a description of the one or more operations and/or the one or more operating engine(s). Optionally, the operations instructions may comprise structured data associated with a structure of operating input data required by the one or more operating engine(s) for generating operating output data. Unstructured data may be for example string data, image data and/or numerical data. Numerical data may be associated with a number and optionally, a corresponding unit.

The operating instructions and the request may be provided to a selection model for generating model output data associated with the selected operation and/or the selected operation engine 208. This may comprise generating model input data by combining the operation instructions and the request for receiving the chemical product with the target property. Followingly, the model input data may comprise unstructured and optionally structured data.

The selection model may be configured to receive the model input data and generating model output data from the model input data. The selection model may be parametrized and/or trained as described within the context of FIG. 18 and FIG. 22. The selection model may be configured to map the model input data to numerical representation of the model input data. Hence, the selection model may comprise one or more embedding layers as described within the context of FIG. 18 and/or one or more encoder inputs 1978, 1988 and/or one or more decoder inputs 1984, 1994. Further, the selection model may be configured to map the numerical representation of the model input data to numerical representation of the model output data by using one or more mathematical relations. Hence, the selection model may comprise one or more mathematical relations. For example, the selection model may comprise one or more encoder blocks 1974, 1986 and/or one or more decoder blocks 1980, 1990 for mapping the numerical representation of the model input data to the numerical representation of the model output data. Further, the numerical representation of the model output data may be mapped to model output data by one or more decoder output 1992, 1982, and/or encoder output 1976. Mapping the numerical representation of the model output data to the model output data may comprise applying one or more mathematical relations, preferably inverse mathematical relations to one mathematical relations used for generating the numerical representation of the model output data from the model input data. By doing so, the unstructured model input data is mapped to a structured representation suitable for being processed by the selection model. This structured representation

requires less computational resources for being processed than the model input data. Further, this allows to process unstructured requests by computing resources requiring structured input.

Further, the model output data may be indicative of the selected operating engine. Hence, the model output data may be suitable for identifying the selected operating engine. Therefore, the selection model may select the selected operating engine for carrying out the selected operation. Where the request may be a request for receiving the chemical product with the target property, the selected operating engine may be configured to provide a digital representation of a chemical structure associated with the chemical product.

In an embodiment, the model output data may be suitable for being received by the at least one selected operating engine. In this embodiment, the operating instructions may further comprise data structures associated with the one or more operating engine(s) comprising at least one data structure associated with the selected operating engine. Hence, the model output data may comprise the one or more data structures. The model output data obtained in response to providing the model input comprising the data structures to the selection model may be suitable for being provided to the selected operating engine for carrying out the selected operation.

The model input data may be a prompt e.g. to a large language model. The selection model may be a large language model. Further data types may be received by the selection model. For this purpose, the selection model may comprise a plurality of different embedding layers such as one or more embedding layers for processing images as described in the context of FIG. 22 and/or one or more embedding layers for processing numerical, in particular tabular data, as described within the context of FIG. 21.

A large language model may be associated with a model architecture as described within the context of FIG. 19A - FIG. 19C.

One or more data structures associated with the one or more operating engine(s) comprising at least one data structure associated with the selected operating engine may be received 212. The data structures may be stored in a data structures. The data structure repository 512 may comprise a database. Hence, the data structures may be retrieved from the data structure repository 512, e.g. by providing a query for receiving the data structures to the data structure repository 512. Preferably, the query for receiving the data structures may be provided to the data structure repository 512 in response to receiving the model output data. The data structures may specify the data structures to be received by the one or more operating engine(s), in particular the data structure to be received by the at least one selected operating engine.

Operating input data may be generated by providing the one or more data structures and the model output data to a structure model configured to generate operating input data from the from the one or more data structures and the model output data 214. The operating input data may be associated with a data structure configured to be received by the selected operating engine for carrying out the selected operation. Hence, the structure model may be configured to extract the data required by the at least one selected operating engine from an unstructured request for receiving the chemical product with the target property for carrying out the at least one selected operation, in particular for generating a chemical structure of the chemical product with the target property. This enables robust and tailored production of chemical products. Ultimately, this results in generating chemical products with a high quality reducing the consumption of materials for otherwise wasted for defective goods.

Specifically, generating operating input data may comprise generating structure input data by combining the model output data and the one or more data structures. The structure input data may be suitable for being received by the structure model. Preferably, the data structures and the model output data may be received together to allow the structure model to select the data structure corresponding to the selected operating engine in relation to the request. By doing so, structured operating input data can be generated for robust providing of chemical product data based on unstructured requests. The structure input data may be a prompt, in particular to a large language model. The structure model may be a large language model.

Further, generating operating input data may comprise providing the structure input data to the structure model for generating the operating input data associated with the data structure suitable for being provided to the at least one selected operating engine.

The structure model may be parametrized and/or trained as described within the context of FIG. 18 and FIG. 22. The structure model may be configured to map the structure input data to numerical representation of the structure input data. Hence, the structure model may comprise one or more embedding layers as described within the context of FIG. 18 and/or one or more encoder inputs 1978, 1988 and/or one or more decoder inputs 1984, 1994. Further, the structure model may be configured to map the numerical representation of the structure input data to numerical representation of the operating input data by using one or more mathematical relations. Hence, the structure model may comprise one or more mathematical relations. For example, the structure model may comprise one or more encoder blocks 1974, 1986, one or more encoder outputs 1976, one or more decoder outputs 1992 and/or one or more decoder blocks 1980, 1990 for mapping the numerical representation of the structure input data to the numerical representation of the operating input data. Further, the numerical representation of the operating input data may be mapped to operating input data by selecting the one or more element(s) associated with the operating input data according to the one or more confidence score(s) associated with the one or more element(s). Mapping the numerical representation of the operating input data to the

operating input data may comprise applying one or more mathematical relations, preferably inverse mathematical relations to one mathematical relations used for generating the numerical representation of the structure input data from the structure input data. By doing so, the unstructured structure input data is mapped to a structured representation suitable for being processed by the structure model. This structured representation requires low computational resources for being processed. Further, this allows to process unstructured requests by computing resources requiring structured input. The so-created operating input data may be suitable for being provided to the selected operating engine for carrying out the selected operation. Further, the structure input data may be indicative of the selected operating engine. Therefore, the structure model may provide and/or generate the data associated with the data structure required by the selected operating engine for carrying out the selected operation. By generating operating input data from the model output data, the data obtained by processing the request may be structured to obtain the data structure required by the selected operating engine. Followingly, operating engine(s) requiring fixed data structures such as databases or purpose-specific data-driven model(s) may be utilized for obtaining the chemical product data. Such operating engine(s) offer highly reliable retrievable of accurate data. Hence, generating operating input data from the model output data enables accurate and robust generation of chemical product data. This saves significant computing resources and allows to process a plurality of requests.

To ensure robust generoutation of operating input data, the data structure related to the operating input data may be validated 218. Validating the data structure related to the operating input data may refer to determining if the data structure related to the operating input data generated by the structure model corresponds to the data structure of input data to the at least one selected operating engine. Validating the data structure related to the operating input data may comprise generating validation input data by combining the operating input data and the received data structures. The validation input data may be provided to the validation model. The validation model may be a large language model analogous to the selection model and/or the structure model. In an embodiment, the validation model may be the same model as the selection model and/or the structure model. This saves resources as less models need to be trained and run. In another embodiment, the validation model may comprise a plurality of deterministic functions for assessing if the data structure related to the operating input data generated by the structure model corresponds to the data structure of input data to the at least one selected operating engine. Hence, the validation model may compare the data structure related to the operating input data with the data structure associated with the selected operating engine. In response to determining that the data structure related to the operating input data corresponds to the data structure of input data to the at least one selected operating engine, the operating input data may be validated. Validating the operating input data may trigger to provide the operating input data to the at least one selected operating engine. Validating the operating input data allows to verify the data structure of the operating input data. This reduces errors in operating the selected operating engine. Hence, computational resources are saved.

The operating input data may be provided to the selected operating engine for generating the operating output associated with the chemical product 220. Where receiving a request for receiving a chemical product with target property was received, the operating engine may be configured to provide and/or generating a digital

5 representation of the chemical structure of the chemical product. The operating engine may be as described within the context of FIG. 6A - FIG. 6F.

The operating output data may be received from the operating engine. The operating output data may be provided together with the request for receiving the chemical product with the target property to a processing model for
10 generating chemical product data comprising at least a part of the operating output data 222. The chemical product data may be indicative of the requested chemical product. The chemical product data may be a response to the request for receiving the chemical product with the target property. Hence, the chemical product data may correspond to the request for receiving the chemical product with the target property.

15 Generating the chemical product data may comprise generating processing task instruction by merging the operating output data and the request for receiving the chemical product with the target property, in particular into one prompt. The processing task instruction may be provided to the processing model. The processing model may be configured to receive the data generation model input data. The processing model may comprise one or more embedding layers as described within the context of FIG. 18 and/or one or more encoder inputs 1978, 1988 and/or
20 one or more decoder inputs 1984, 1994. The processing model, in particular the embedding layers of the processing model, may be configured to map the processing task instruction to numerical representation of the processing task instruction. Further, the processing model may be configured to map the numerical representation of the processing task instruction to numerical representation of the chemical product data by using one or more mathematical relations. Hence, the processing model may comprise one or more mathematical relations. For
25 example, the processing model may comprise one or more encoder blocks 1974, 1986 and/or one or more decoder blocks 1980, 1990 for mapping the numerical representation of the processing task instruction to the numerical representation of the chemical product data. Further, the numerical representation of the model chemical product data may be mapped to chemical product data by one or more decoder output 1992, 1982, and/or encoder output 1976. Mapping the numerical representation of the chemical product data to the chemical
30 product data may comprise applying one or more mathematical relations, preferably inverse mathematical relations to one mathematical relations used for generating the numerical representation of the chemical product data from the processing task instruction. By doing so, the unstructured processing task instruction is mapped to a structured representation suitable for being processed by the processing model. This structured representation requires low computational resources for being processed. Further, this allows to process unstructured requests
35 by computing resources requiring structured input. In an embodiment, the processing model may be the same

model as the selection model and/or the structure model and/or the validation model. This saves resources as less models need to be trained and run.

The generated chemical product data may be provided 224 e.g. via a user interface and/or to an equipment interface 120 as described within the context of FIG. 1. Hence, the chemical product data may be suitable for being provided to a control unit configured to control equipment of a chemical production facility. The chemical product data may comprise control data suitable for controlling a control unit configured to control equipment of a chemical production facility.

FIG. 3 illustrates an embodiment of a method for obtaining chemical product data associated with a chemical product.

A request for receiving chemical product data associated with a chemical product may be received 302. The request may be received as described within the context of FIG. 2.

The request may indicate a chemical product. The request may comprise a digital representation of the chemical product and/or one or more properties associated with the chemical product. For example, the digital representation of the chemical product may be a denotation of the chemical product and/or one or more components of the chemical product. The denotations may include for example trivial name, commercial name, IUPAC name, SMILES, SMARTS or the like. Further, the digital representation of the chemical product may be an image of the chemical product and/or measurement results obtained by analyzing the chemical product, e.g. by spectroscopical methods such as infrared spectroscopy.

Operating instructions associated with one or more operations carried out by one or more operating engine(s) may be received 306 analogous to 206. The one or more operations may comprise a selected operation related to the request. The one or more operating engines may comprise a selected operating engine configured to carry out the selected operation instructions.

The operating instructions and the request for receiving the chemical product data may be provided to a selection model 308 as described within the context of 208.

One or more data structures associated with the one or more operations comprising at least one data structure associated with the selected operating engine may be received 312 analogous to 212.

Operating input data may be generated by providing the one or more data structures and the model output data to a structure model configured to generate operating input data from the from the one or more data structures and the model output data 314 analogous to 214. The operating input data may be associated with a data structure configured to be received by the selected operating engine for carrying out the selected operation.

5

The data structure related to the operating input data may be validated by providing the one or more data structures and the operating input data to a validation model 318 analogous to 218.

Operating output data may be generated by providing the operating input data to the selected operating engine 320 analogous to 220. The operating output data may be associated with the chemical product and/or a property of the chemical product.

Chemical product data associated with the chemical product and/or the property of the chemical product may be generated by providing the request and the operating output to a processing model configured to generate chemical product data from the request and the operating output data 322 analogous to 222. The chemical product data may comprise at least a part of the operating output data.

Where the request for receiving the chemical product data comprises image data, the models for processing data associated with the request for receiving the chemical product data may comprise one or more embedding layers as described in the context of FIG. 22. Where the request for receiving the chemical product data comprises numerical data such as tabular data, the models for processing data associated with the request for receiving the chemical product data may comprise one or more embedding layers as described in the context of FIG. 21. Where the request for receiving the chemical product data comprises text data, the models for processing data associated with the request for receiving the chemical product data may comprise one or more embedding layers as described in the context of FIG. 18. The models for processing the request for receiving the chemical product data may be for example the selection model, the structure model, the validation model, one or more models of the execution executing service 6-114, the processing model or the like.

The chemical product data may be provided 324 analogous to 224.

30

FIG. 4 illustrates an embodiment of a method for obtaining a digital representation of a chemical product.

Chemical structures are defined and complex structures. Thus, highly regulated nomenclatures are required for describing a chemical product. The thalidomide scandal is an alarming example of how an orientation of a functional group of a chemical product may result in completely different behaviour and thus, properties of a

35

chemical product. Similarly, chemical production is highly dependent on the chemical identity of the chemical products to deliver robust and tailored chemical products for further processing toward end-of-life products. Hence, the clear distinction between chemical products is of high technical relevance to ensure robust production and processing of chemical products. Obtaining a digital representation of a chemical product by providing a request to a selection model and/or a selection engine enables an efficient and robust relation from an unstructured request to structured data. By doing so, exact digital representations of chemical products related to the request can be obtained.

A request for receiving a digital representation of a chemical product may be received 402 analogous to 202 and/or 302.

The request may indicate a chemical product. The request may comprise a digital representation of the chemical product and/or one or more properties associated with the chemical product. For example, the digital representation of the chemical product may be a denotation of the chemical product and/or one or more components of the chemical product. The denotations may include for example trivial name, commercial name, IUPAC name, SMILES, SMARTS or the like. Further, the digital representation of the chemical product may be an image of the chemical product and/or measurement results obtained by analyzing the chemical product, e.g. by spectroscopical methods such as infrared spectroscopy. Further, the request may indicate a type of the chemical product such as a polymer, an inorganic salt or the like and/or a field of application associated with the chemical product.

Operating instructions associated with one or more operations carried out by one or more operating engine(s) may be received 406 analogous to 206 and/or 306. The one or more operations may comprise a selected operation related to the request. The one or more operating engines may comprise a selected operating engine configured to carry out the selected operation instructions.

The operating instructions and the request for receiving the chemical product data may be provided to a selection model 408 analogous to 208 and/or 308. The selection model may determine if the request may be sufficient for identifying the chemical product associated the receiving a request for receiving the digital representation of a chemical product

One or more data structures associated with the one or more operations comprising at least one data structure associated with the selected operating engine may be received 412 analogous to 212 and/or 312.

Operating input data may be generated by providing the one or more data structures and the model output data to a structure model configured to generate operating input data from the from the one or more data structures and the model output data 414 analogous to 214 and/or 314. The operating input data may be associated with a data structure configured to be received by the selected operating engine for carrying out the selected operation.

5

The data structure related to the operating input data may be validated by providing the one or more data structures and the operating input data to a validation model 418 analogous to 218 and/or 318.

Operating output data may be generated by providing the operating input data to the selected operating engine 420 analogous to 220 and/or 320. The operating output data may be associated with the chemical product and/or a property of the chemical product.

Chemical product data associated with the chemical product and/or the property of the chemical product may be generated by providing the request and the operating output to a processing model configured to generate chemical product data from the request and the operating output data 422 analogous to 222 and/or 322. The chemical product data may comprise at least a part of the operating output data.

Where the request for receiving the chemical product data comprises image data, the models for processing data associated with the request for receiving the chemical product data may comprise one or more embedding layers as described in the context of FIG. 22. Where the request for receiving the chemical product data comprises numerical data such as tabular data, the models for processing data associated with the request for receiving the chemical product data may comprise one or more embedding layers as described in the context of FIG. 21. Where the request for receiving the chemical product data comprises text data, the models for processing data associated with the request for receiving the chemical product data may comprise one or more embedding layers as described in the context of FIG. 18. The models for processing the request for receiving the chemical product data may be for example the selection model, the structure model, the validation model, one or more models of the execution executing service 6-114, the processing model or the like.

The chemical product data may be provided 424 analogous to 224 and/or 324.

30

FIG. 5 illustrates an embodiment of an operating system 538.

The operating system 538 may comprise a system for obtaining a chemical product, a system for obtaining chemical product data associated with a chemical product and/or a system for obtaining a digital representation associated with a chemical product. The operating system 538 may comprise an intake interface 116, a selection

35

service 104, a structure service 110, a validation service, a chemical data generating service 112 and/or an output interface 118.

5 The intake interface 116 may be suitable for receiving a request for receiving the chemical product data, a request for receiving the chemical product with the target property and/or a request for receiving a digital representation of a chemical product. The intake interface may for example a user interface. The intake interface 116 may allow a user to interact with the operating system 538 e.g. for obtaining a chemical product, for obtaining chemical product data associated with a chemical product and/or obtaining a digital representation associated with a chemical product. The intake interface 116 may be configured to receive one or more requests according to 202, 302 and/or
10 402.

Operation instructions may be received from an operation instructions repository 506 as described within the context of 206, 306 and/or 406 in FIG. 2 - FIG. 4.

15 The one or more requests may be provided to the selection service 104, in particular the selection processing engine 504. The selection service may comprise the operation instructions repository 506, the selection processing engine 504 and/or the selection engine 508. The selection processing engine 504 may be configured to generate the model input data by combining the operation instructions and the request for receiving the chemical product data, a request for receiving the chemical product with the target property and/or a request for
20 receiving a digital representation of a chemical product as described within the context of FIG. 2 - FIG. 4. The model input data may be provided to the selection engine 508 for generating model output data according to 208, 308 and/or 408 as described within the context of FIG. 2 - FIG. 4. Hence, the selection engine may comprise the selection model or may interface the selection model e.g. via an API. Where the selection model may interface the selection model, the selection model may be configured to provide the model input data to the selection model
25 and receiving the model output data from the selection model. In an example, the selection model may be hosted by a different entity than an entity associated with the operating system 538.

The model output data may be provided to the structure service 110. The structure service 110 may comprise a structure processing engine 514, a structure engine 510 and/or a data structure repository 512. The data structure
30 repository 512 may store the data structures associated with the one or more operations comprising at least one data structure associated with the selected operating engine. The data structures may be received from the data structure repository 512 as described within the context of 212 in FIG. 2. Hence, the data structure repository may be configured to provide the data structures. The structure processing engine 514 may be configured to combine the data structures and the model input data as described for 214, 314 and/or 414 in the context of FIG. 2 - FIG.
35 4. The operating input data may be provided from the structure processing engine 514 to the structure engine for

generating the operating input data. The structure engine 510 may be configured to generate the operating input data according to 214, 314 and/or 414. Hence, the structure engine 510 may comprise the structure model as described within the context of 208, 308 and/or 408. Alternatively, the structure engine 510 may interface the structure model. Hence, the structure model may be configured to call an API towards to structure model.

- 5 Followingly, the structure model may be configured, in particular trained and/or parametrized, analogous to the selection model.

The operating input data may be provided to the validation service 106, in particular from the structure engine 510. The validation service 106 may comprise a validation processing engine 516, a validation engine 518 and/or
10 a data structure repository 548. Alternatively, the validation service may be in connection with the data structure repository 512 associated with the structure service 110. In an embodiment, the structure service 110 may comprise the structure engine 510, the structure processing engine 514, the data structure repository 512, the validation processing model and the validation engine. The validation processing engine 516 may be configured to combine the data structures and the operating input data according to 218, 318 and/or 418. The data structures
15 may be received from the data structure repository 512 or the data structure repository 548 according to 218, 318 and/or 418. The operating input data may be provided from the validation processing engine 516 to the validation engine 518. The validation engine may be configured to validate the operating input data according to 218, 318 and/or 418. Hence, the validation engine 518 may comprise the validation model as described within the context of 218, 318 and/or 418. Alternatively, the validation engine 518 may interface the validation model. Hence, the
20 validation engine 518 may be configured to call an API towards the validation model. Followingly, the validation model may be configured, in particular trained and/or parametrized, analogous to the selection model and/or the structure model.

The operating input data may be provided to the executing service 108, in particular the selected operating
25 engine, preferably in response to validating the operating input data according to 220, 320 and/or 420. The executing service 108 may comprise one or more operating engine(s) comprising the selected operating engine 550. At least one of the one or more operating engine(s) comprising the selected operating engine 550 ay be the selected operating engine. Examples of operating engine(s) are described in the context of FIG. 6A - FIG. 6F. From the selected operating engine operating output data may be received as described within the context of 220,
30 320 and/or 420.

The operating output data may be provided from the selected operating engine to the chemical data generating service 112, in particular the output processing engine 552. The chemical data generating service 112 may
35 comprise the chemical data generating engine 522 and/or the output processing engine 552. The output processing engine 552 may receive the operating output data from the selected operating engine. The output

processing engine 552 may receive the request for receiving the chemical product data, the request for receiving the chemical product with the target property and/or the request for receiving a digital representation of a chemical product from the intake interface 116. The output processing engine 552 may be configured to generate processing task instruction from the operating output data and at least one of the request for receiving the chemical product data, the request for receiving the chemical product with the target property and/or the request for receiving a digital representation of a chemical product according to 222, 322 and/or 422 as described within the context of FIG. 2 - FIG. 4. The processing task instruction may be provided from the output processing engine 552 to the chemical data generating engine 554. The chemical data generating engine 552 may be configured to generate chemical product data from the processing task instruction according to 222, 322 and/or 422 as described within the context of FIG. 2 - FIG. 4.

The chemical product data may be provided by the output interface 118 as described within the context of FIG. 1 and/or according to 224, 324 and/or 424 as described within the context of FIG. 2 - FIG. 4.

FIG. 6A illustrates an embodiment of an operating engine and an executing service 108.

The executing service 108 may comprise one or more operating engine(s) as described in the context of FIG. 5. In an example, at least one operating engine may be a structured database 650. The structured database 650 may receive operating input data from a structure service and/or validation service 630. The structure service 630 may correspond to the structure service 110 as described in the context of FIG. 1. The validation service 630 may correspond to the validation service 106 as described in the context of FIG. 1. The structured database 650 may be configured to receive a query from the structure service and/or validation service 630. Hence, the operating input data may comprise the query where the selected operating engine may be the structured database 650. The query may be associated with a predefined data structure. Further, the structured database 650 may be configured to retrieve operating output data in response to receiving the query. The structured database 650 may comprise predefined relations between one or more chemical product data sets. Retrieving the operating output data may comprise selecting at least a part of the chemical product data sets corresponding to the query. The query may be indicative of at least the part of the chemical product data sets. In an example, the structured database 650 may be a SQL database. This ensures reliable retrieval of the operating output data. The operating output data retrieved by the structured database 650 may be received by the chemical data generating service 644. The chemical data generating service 644 may correspond to the chemical data generating service 112 as described in the context of FIG. 1.

FIG. 6B illustrates an embodiment of an operating engine and an executing service 108.

The executing service 108 may comprise one or more operating engine(s) as described in the context of FIG. 5. In an example, at least one operating engine may be an embedding database 646. The embedding database 646 may receive operating input data from a structure service and/or validation service 630. The structure service 630 may correspond to the structure service 110 as described in the context of FIG. 1. The validation service 630 may correspond to the validation service 106 as described in the context of FIG. 1. The operating input data may be mapped to an embedded operating input data. The embedded operating input data may comprise a numerical representation of the operating input data. For example the embedded operating input data may be a tensor, in particular a vector. An example for the embedded operating input data may be the embedded input 1814. The embedded operating input data may be obtained by passing the operating input data through one or more embedding layers 1802. An example of an embedding layer and obtaining the embedding layer may be described in FIG. 18. The embedding database 646 may comprise a plurality of chemical product data sets. Representations of the chemical product data sets may be obtained analogous to the representation of the operating input data. Similarly, the representations of the chemical product data sets may be embedded chemical product data sets. Retrieving the operating output data from the embedding database 646 may comprise selecting at least a part of the chemical product data sets by determining if the distance between the embedded operating input data and the embedded chemical product data may be within a predefined range. The distance between the embedded operating input data and the embedded chemical product data set may be an euclidean distance and/or a cosine distance between the embedded operating input data and the embedded chemical product data. In an embodiment, the chemical product data set associated with a smaller distance between the embedded operating input data and the embedded chemical product data set than the distance between the embedded operating input data and the other embedded chemical product data sets may be selected. This may be advantageous since the operating output data may be retrieved accurately even if the operating input data may comprise for example string data. Different words for describing the same matter may be available. An embedding database 646 can relate different words with the same meaning. Chemical products may be associated with a plurality of different nomenclatures such as trivial names or IUPAC names. Hence, using embedding databases 646 for obtaining the operating output data provides accurate retrieval even if chemical product data sets from different domains may be available. This saves resources for harmonizing the documentation associated with chemical product data sets. The operating output data retrieved by the embedding database 646 may be received by the chemical data generating service 644. The chemical data generating service 644 may correspond to the chemical data generating service 112 as described in the context of FIG. 1.

FIG. 6C illustrates an embodiment of an operating engine and an executing service 108.

The executing service 108 may comprise one or more operating engine(s) as described in the context of FIG. 5. In an embodiment, the operating engine may be a data-driven model 652. The data-driven model 652 may be

configured to receive the operating input data from the structure service and/or validation service 630. The data-driven model may generate operating output data from the operating input data. The operating output data may be received by the chemical data generating service 644. The chemical data generating service 644 may correspond to the chemical data generating service 112 as described in the context of FIG. 1. The data-driven model may

5 comprise one or more mathematical equations associated with a relation between the operating input data and the operating output data. In an embodiment, the data-driven model may be a neural network. The neural network may comprise a plurality of neurons. A neuron may describe a mathematical relation between its input and its output. The neural network may comprise one or more input layers, one or more hidden layers and/or one or more output layers. The input layer(s) may be configured to receive the operating input data. The operating input data

10 may be associated with a data structure suitable for being received by the input layer(s). The input layer(s) may comprise a plurality of neurons. The neurons of the input layer(s) may be connected to the neurons of the hidden layer(s). Hence, the output of a neuron of the input layer(s) may be provided to a neuron of the hidden layer(s). The neurons of the hidden layer(s) may be connected to the neurons of the output layer(s). Hence, the output of a neuron of the hidden layer(s) may be provided to a neuron of the output layer(s). The output layer(s) may output

15 the operating output data. The data-driven model may be suitable for describing non-linear relations between datapoints. Hence, using a data-driven model as an operating engine may allow to obtain operating output data where no measurement data may be available. This saves a high amount of resources otherwise needed to obtain the measurement data.

20 FIG. 6D illustrates an embodiment of an operating engine and an executing service 108.

The executing service 108 may comprise one or more operating engine(s) as described in the context of FIG. 5. In an embodiment, the operating engine may be a user interface 654. The user interface may be configured to receive the operating output data in response to providing the operating input data. The operating output data may

25 be received by the chemical data generating service 644. The chemical data generating service 644 may correspond to the chemical data generating service 112 as described in the context of FIG. 1. The operating input data may be provided by the structure service and/or validation service 630. The user interface may receive the operating output data from a user. Hence, the user interface 654 allows a user to interact with the operating system. This is beneficial in cases where the user is a domain expert. Chemical production has high safety

30 requirements and thus, requires transparent decisions when for example controlling a chemical production facility 122. Using a user interface as an operating engine enables an efficient human-machine interaction. Further, the additional information for obtaining the chemical product, the representation of the chemical product and/or the chemical product data may be received by the user interface. Therefore, requests may be enhanced which allows for an efficient processing of the request.

FIG. 6E illustrates an embodiment of an operating engine and an executing service 108.

The executing service 108 may comprise one or more operating engine(s) as described in the context of FIG. 5. In an embodiment, the operating engine may comprise a selection service 104, a structure service 110, a validation service 106 and/or an executing service 108. The selection service 104, the structure service 110, the validation service 106 and/or the executing service 108 may be as described in the context of FIG. 1. The selection service 104 may receive the operating input data from the structure service and/or validation service 630. The selection service 104 may provide the processed operating input data to the structure service 110. The structure service 110 may structure the processed operating input data. The structured operating input data may be provided to the validation service 106 for validating the data structure associated with the structured operating input data. The structured operating input data may be provided to the executing service 108 for generating the operating output data. The operating output data may be received by the chemical data generating service 644. The chemical data generating service 644 may correspond to the chemical data generating service 112 as described in the context of FIG. 1. The operating input data may be provided by the structure service and/or validation service 630. This may allow tailor the selection by the selection model conducted by a further selection. For example, the selection service 104 may select a group of operating engines. Then, a further selection may be beneficial to further select one or more operating engine(s) from the group of operating engine(s). This may be particularly advantageous if the request received may be insufficient for determining a selected operating engine. Additional information may be received e.g. via a user interface as described in the context of FIG. 6D. The further selection may be conducted by the selection service 104 based on the request and the additional information. Hence, the executing service 108 as described in FIG. 6E may allow for efficient processing of the request to save computational resources.

FIG. 6F illustrates an embodiment of an operating engine and an executing service 108.

The executing service 108 may comprise one or more operating engine(s) as described in the context of FIG. 5. In an embodiment, the operating engine may comprise an intake interface 6-124, a chemical structure generating engine 6-132, a compound database 6-126 comprising digital representations of a chemical structure of one or more educts 6-128, a property determining engine 6-134, a formation score determining engine 6-130, a product determining engine 6-136 and/or an output interface 6-138. The intake interface 6-124 may be configured to receive the operating input data from the structure service and/or validation service 630. The operating input data may be provided by the structure service and/or validation service 630. The structure service 630 may correspond to the structure service 110 as described in the context of FIG. 1. The validation service 630 may correspond to the validation service 106 as described in the context of FIG. 1. The intake interface 6-124 may provide the operating input data to the chemical structure generating engine 6-132. The chemical structure generating engine

6-132 may be configured to generate a digital representation of the chemical product from the operating input data. The operating input data may comprise a target property of the chemical product. The chemical structure generating engine 6-132 may determine digital representations of chemical products obtained by a chemical reaction of one or more educts from digital representations of a chemical structure of one or more educts 6-128.

- 5 For this purpose, the chemical structure generating engine 6-132 may receive digital representations of a chemical structure of one or more educts 6-128 from the compound database 6-126. A request for receiving the digital representations of a chemical structure of one or more educts 6-128 may be provided by the chemical structure generating engine 6-132 to the compound database 6-126. The request for receiving the digital representations of a chemical structure of one or more educts 6-128 may be a query suitable for being received by
- 10 the compound database 6-126. The compound database 6-126 may be a structured database as described in the context of FIG. 6A.

- The digital representations of chemical structures of chemical products may be provided to a formation score determining engine 6-130. The formation score determining engine 6-130 may be configured to determine a
- 15 formation score associated with a formation of the chemical products from the one or more educts. For example, a high formation score may indicate a high rate of formation of the chemical products. The formation score may be obtained by calculating the degree of atomic configurations unchanged by the chemical reaction of the one or more educts to the chemical products. As the production of chemical products may be associated with equilibria and incomplete conversions, the formation score may be indicative of the efficiency of a production process. The
- 20 chemical product may be selected if the formation score associated with the formation of the chemical product may be within a predefined range. This allows for increasing the efficiency of the production of chemical products and reduces the amount of undesired byproducts.

- Further, the digital representations of chemical structures of the chemical products may be received by the
- 25 property determining engine 6-134. The property determining engine 6-134 may be configured to determine a property of the chemical product from the digital representations of the chemical products. For example, the property determining engine 6-134 may comprise a classification model.

- The determined properties may be provided to a product determining engine 6-136. The determined formation
- 30 score may be provided to the product determining engine 6-136. The digital representations of a chemical structure of the chemical products may be provided to the product determining engine 6-136. The product determining engine 6-136 may be configured to select the chemical product associated with the target property e.g. by comparing the properties of the chemical product with the target property. The product determining engine 6-136 may further select the chemical product by determining that the formation score may be within a predefined
- 35 range. Hence, the property determining engine 6-134 may select the chemical product with the target property

from the chemical products associated with the digital representations of the chemical structure of the chemical products as obtained by the chemical structure generating engine 6-132. The product determining engine 6-136 may provide the digital representation of the chemical structure to the output interface 6-138. Hence, the operating output data may be associated with a digital representation of the chemical structure of the chemical product
5 associated with the target property.

By using the above described engines and/or conducting the above described acts, the chemical product with the target property can be obtained efficiently from e.g. educts available to the chemical production facility 122.

10 The operating output data may be received by the chemical data generating service 644. The chemical data generating service 644 may correspond to the chemical data generating service 112 as described in the context of FIG. 1.

FIG. 7 illustrates an embodiment of producing and/or processing a chemical product 714.

15 The chemical product may be produced by the chemical production facility 702. The chemical production facility 702 may be in connection to an operating system of a chemical production facility 102. Hence, the chemical production facility 702 may be monitored and/or controlled by the operating system of a chemical production facility 102. The operating system of a chemical production facility 102 may be as described in the context of FIG.

20 1. The chemical product may be provided from the chemical production facility 702 to a chemical product processing facility 704. The chemical product processing facility 704 may be controlled and/or monitored by an operating system of a chemical product processing facility 706. The operating system of a chemical product processing facility 706 may comprise an intake interface 712, a request providing service 710 and/or an output interface 708. For processing of the chemical product 714, the chemical product 714 may be associated with
25 target properties. The target properties may be prescribed by the chemical product processing facility 704 and/or may be a result of the target processing of the chemical product. Hence, processing specifications may be provided to the intake interface 712. The intake interface 712 may be configured to receive the processing specifications. The intake interface 712 may provide the processing specifications to the request providing service 710. The request providing service 710 may be configured to generate a request for receiving a chemical product
30 associated with the target property from the processing specifications. The request may be provided to the output interface 708. The output interface 708 may provide the request to the intake interface 116. The request may be processed as described within the context of FIG. 2 - FIG. 4.

FIG. 8 illustrates an embodiment of producing and/or processing a chemical product 814.

35

The chemical product may be produced by the chemical production facility 802. The chemical production facility 802 may be in connection to an operating system of a chemical production facility 102. Hence, the chemical production facility 802 may be monitored and/or controlled by the operating system of a chemical production facility 102. The operating system of a chemical production facility 102 may be as described in the context of FIG.

- 5 1. The chemical product may be provided from the chemical production facility 802 to a chemical product processing facility 804. The chemical product processing facility 804 may be controlled and/or monitored by an operating system of a chemical product processing facility 804. The operating system of a chemical product processing facility 804 may comprise an intake interface 812, a request providing service 810, an equipment interface 824 and/or an output interface 808.

- 10 The request providing service 810 may generate a request for receiving chemical product data e.g. for adapting the chemical product processing facility 804 according to the chemical product received and/or properties of the chemical product. Inadequate treatment of chemical products significantly decrease the performance of the chemical product during processing and/or application. Hence, the chemical product processing facility 804 may
- 15 require to be controlled according to the chemical product received from the chemical production facility 122. For this purpose, the request generated by the request providing service 710 may be provided to the output interface 808. The output interface 808 may provide the request to the intake interface 116 of the operating system of a chemical production facility 102. The operating system of a chemical production facility 102 may process the request as described in the context of FIG. 1 - FIG. 4. The output interface 118 of the operating system of a
- 20 chemical production facility 102 may provide the chemical product data to the intake interface 812 of the operating system of a chemical product processing facility 804. Further, the chemical product data may be provided by the intake interface 812 to the equipment interface 824. The equipment interface 824 may be configured to control the chemical product processing facility 804. Hence, the processing to the chemical product can be improved by retrieving chemical product data. Advantageously, the request may comprise unstructured data while the chemical
- 25 product data can be structured and hence, machine-readable. By doing so, the request providing service 710 may comprise a user interface e.g. for entering string data while the chemical product processing facility 804 can be controlled by structured control data obtained from the chemical product data.

- FIG. 9 illustrates an embodiment of the input and output data associated with the selection model, the structure
- 30 model, the validation model, the one or more operating engine(s) and/or the processing model.

- The selection model may be configured to receive the selection task instruction including the request and the functional specification data. The selection model may generate the model output data from the selection task instruction. An example of input and output data associated with the selection model may be seen in FIG. 10. The
- 35 model output data may be merged with the one or more input data structure(s) to the structure task instruction.

The structure model may be configured to generate the operating input data from the structure task instruction. An example of input and output data associated with the structure model may be seen in FIG. 11. The operating input data may be merged with the one or more input data structure(s) and the indication on the at least one selected operating engine to the validation task instruction. The validation model may be configured to generate the validation indication from the validation task instruction. An example of input and output data associated with the validation model may be seen in FIG. 12. The validation indication may be the indication on whether the data structure related to the operating input data corresponds to the input data structure related to the at least one selected operating engine. If the data structure associated with the operating input data may be validated, the operating input data may be provided to the one or more selected operating engine(s) for generating the operating output data. An example of at least one selected operating engine may be described in the context of FIG. 6F. The operating output data may be merged with the request to the processing task instruction. The processing model may be configured to generate the chemical product data from the processing task instruction. An example of input and output data associated with the processing model may be seen in FIG. 13.

FIG. 10 illustrates an embodiment of the input and output data associated with the selection model.

FIG. 11 illustrates an embodiment of the input and output data associated with the structure model.

FIG. 12 illustrates an embodiment of the input and output data associated with the validation model.

FIG. 13 illustrates an embodiment of the input and output data associated with the processing model.

FIG. 14 illustrates a user interface 1402 for receiving a chemical product with a target property.

The request for receiving the chemical product with the target property may comprise a target property, a target type of the chemical product, a target field of application associated with the chemical product and a description of requested service. In the example, the description of the requested service may comprise unstructured text data. The description of the requested service may be predefined and/or may be specified by a user. The target property, the target type and/or the target field of application may be entered into the user interface e.g. by selection the target values from a plurality of values or by specifying free text in relation to the target property, the target type and/or the target field of application. For this purpose, the user interface 1402 may provide corresponding input fields 1408, 1404, 1410 and/or 1412. This may be depicted schematically in the upper user interface 1402. Once the data may be entered into the user interface 1402, the user interface 1402 may show the entered data in the corresponding fields as depicted in the lower user interface 1402 and an offer to produce the requested chemical product.

FIG. 15 illustrates an embodiment of a user interface 1512 for receiving chemical product data.

The request for obtaining chemical product data may be received and/or provided by the user interface 1512. The request may comprise a name of the chemical product, a type of the chemical product, a field of application associated with the chemical product and/or a description of requested service. In the example, the description of the requested service may comprise unstructured text data. The description of the requested service may be predefined and/or may be specified by a user. The name of the chemical product, the type of the chemical product and/or the field of application of the chemical product may be entered into the user interface e.g. by selection the target values from a plurality of values or by specifying free text in relation to the target property, the target type and/or the target field of application. For this purpose, the user interface 1502 may provide corresponding input fields 1504, 1506, 1508 and/or 1510. This may be depicted schematically in the upper user interface 1512. Once the data may be entered into the user interface 1512, the user interface 1512 may show the entered data in the corresponding fields as depicted in the lower user interface 1512 and the requested chemical product data.

FIG. 16 illustrates an embodiment of a user interface 1612 for receiving a digital representation of a chemical product.

The request for obtaining chemical product data may be received and/or provided by the user interface 1612. The request may comprise a name of the chemical product, a type of the chemical product, a field of application associated with the chemical product, one or more ingredients of the chemical product and/or a description of requested service. In the example, the description of the requested service may comprise unstructured text data. The description of the requested service may be predefined and/or may be specified by a user. The name of the chemical product, the type of the chemical product, the one or more ingredients and/or the field of application of the chemical product may be entered into the user interface e.g. by selection the target values from a plurality of values or by specifying free text in relation to the target property, the target type, the one or more ingredients and/or the target field of application. For this purpose, the user interface 1612 may provide corresponding input fields 1604, 1606, 1608 1624 and/or 1610. This may be depicted schematically in the upper user interface 1612. Once the data may be entered into the user interface 1612, the user interface 1612 may show the entered data in the corresponding fields as depicted in the lower user interface 1612 and the requested digital representation of the chemical product. In the example, the requested digital representation of the chemical product Glysantin may be a list of ingredients. This may be required where processing of a specific chemical should be prevented. Even small amounts of a chemical product may have a large influence on the processing of the chemical product. For example, small amounts of metal ions as attached to a stirring bar even after thorough cleaning may catalyse unwanted reactions and thus, change the properties of the chemical product. Hence, the correct digital representation of a chemical product is of high importance to ensure an efficient processing of chemical products.

FIG. 17 illustrates embodiments of APIs for obtaining a chemical product with a target property, chemical product data associated with a chemical product and/or a digital representation of a chemical product.

- 5 For obtaining a chemical product with a target property, chemical product data associated with a chemical product and/or a digital representation of a chemical product a selection model may be deployed. The selection model may be described in more detail in 208, 308 and/or 408. The selection model may be called via a selection model API 1708. The selection model API 1708 may be configured to receive operating instructions and a request for obtaining a chemical product with a target property, chemical product data associated with a chemical product and/or a digital representation of a chemical product, in particular model input data. Further, the selection model API 1708 may be configured to receive the model output data from the selection model.

- The model output data and data structures associated with the one or more operating engine(s), in particular structure input data, may be received by the structure model API 1710. Further, the structure model API 1710 may provide the structure input data to a structure model as described within the context of 214, 314 and/or 414. Further, the structure model API 1710 may be configured to receive the operating input data from the structure model.

- The operating input data and the data structures may be provided to a validation model via the validation model API 1712. Further, the validation model API 1712 may be configured to receive the validated operating input data from the validation model. The validation model may be as described within the context of 218, 318 and/or 418.

- The operating input data may be received by the operating engine API 1714 and provided to an operating engine as described in the context of 220, 320 and/or 420. The operating engine API 1714 may be configured to receive operating output data from the operating engine.

- The operating output data and the request for obtaining a chemical product with a target property, chemical product data associated with a chemical product and/or a digital representation of a chemical product, in particular the processing task instruction, may be provided to a processing model API 1716. The 1716 may be configured to receive the processing task instruction and providing the processing task instruction to the processing model. The processing model may be as described within the context of 222, 322 and/or 422. Further, the 1716 may be configured to receive the chemical product data from the processing model.

- The equipment interface 120 as described within the context of FIG. 1, FIG. 7 and/or FIG. 8 may comprise an equipment API 1718. The equipment API 1718 may be configured to receive the chemical product data e.g. from

an output interface 118 and providing the chemical product data to equipment of a chemical production facility 122.

FIG. 18 illustrates an embodiment of obtaining an embedding layer. The embedding layer may be obtained by training for example a continuous bag of words model (CBOW) or a skip-gram model. The embedding layer may be suitable for generating embedded input data based on input data. Generating embedded input data may refer to embedding input data. The embedding layer may map data to numerical representation of the data. Embedded data may be used synonymously to numerical representation of the data. Embedding input data may result in a representation associated with the input data. Thus, the embedded input 1814 may be the representation associated with the input data. The input data may comprise one or more elements. The one or more elements may be represented by the input vector 1806. In particular, the embedded input 1814 and/or the input vector 1806 may be machine- readable and/or processable by a processor. For this purpose, the embedded input 1814 and/or the input vector 1806 may be a tensor, in particular a first-rank tensor. Specifically, the input vector 1806 may be a one-hot vector or a summation of a plurality of one-hot vectors. A one-hot vector may be a vector with one entry unequal to zero. Examples for one-hot vectors may be 1808, 1810 and 1812. The entries unequal to zero in the one-hot vector and/or in the input vector 1806 may indicate the element. For example, a lookup table may define the relation between the position of the entries unequal to zero and the element indicated by the one-hot vector. The lookup table may specify a plurality of different elements. The number of different elements may be equal to the number of entries in the one-hot vector. The number of different elements may be referred to as vocabulary size. In an example, the elements may be represented by tokens and a sequence of elements may refer to at least a part of a sentence. The at least a part of the sentence may be represented by a plurality of tokens. A token may represent at least a part of the element and/or word. For example, where one element would be associated with only one word, words such as "embeddings", "embedding" or "embed" would constitute different elements. A first token may represent the stem "embed" and the endings, typically appearing in a plurality of word, may be represented by a second token, a third token and a fourth token. The second token, the third token and the fourth token may be used for representing other words such as "look", "looking" or the like, preferably together with a fifth token representing the stem "look". Ultimately, this tokenization of elements associated with a plurality of stems and a plurality of endings results in less tokens to be used for representing a plurality of elements and thus, uses less computational resources.

A lookup table specifying a subset of the vocabulary size e.g. of the English language may comprise 10,000 words or more. The embedded input 1814 may be a lower-dimensional representation than the input vector 1806. For example, typical embedded inputs 1814 may comprise some hundreds of different entries. Followingly, the embedded inputs 1814 constitute a densified representation of one or more elements using less computational resources. More than that, the embedded input 1814 may represent a relation between two or more elements. For example, the words "Italy" and "Germany" may be similar or may be more closely related since they both define

European countries, whereas the word “embodiment” may be very different from the two respective words. The smaller the dot product between two embedded inputs 1814 may be the more similar the two elements associated with the embedded inputs 1814 may be. Hence, the embedded inputs 1814 may represent one or more elements accurately and lead to accurate results based on processing the embedded inputs 1814.

5 For transforming the input vector 1806 into the embedded input 1814, the embedding layer may comprise a number of neurons equal to the number of entries in the embedded input 1814. Based on the embedded inputs 1814, the output layer may generate the output vector 1816. The output vector may be a vector and/or may indicate one or more elements. The output vector 1816 may indicate one or more elements different from the input vector 1806 and/or the one-hot vectors associated with the input vector 1806. For this purpose, the output layer
10 may comprise a number of neurons equal to the number of entries of the input vector 1806 and/or the output vector 1816. The output layer may apply a softmax function to the embedded inputs 1814. By doing so, the output vector may comprise the probabilities associated with the elements associated with the entries of the output vector 1816 unequal to zero. Hence, from the output vector 1816 one or more elements may be obtained with a corresponding probability. Where the input vector 1806 may specify one or more sequence(s) of elements, the
15 output vector 1816 may specify one or more elements corresponding to the sequence(s) of elements specified by the input vector 1806. In the example of FIG. 18, the element associated with vector 1818 may correspond to the input vector with a probability of 71 %. Additional or alternative elements may correspond to the input vector as indicated by the output vector with lower probability. By defining a threshold to which the probability may be compared, the selection of the corresponding elements may be tailored to the needs of the user. The elements
20 generated by the model comprising the embedding layer 1802 and the output layer 1804 may refer to the most probable elements indicated by the output vector 1816. Hence, the model depicted in FIG. 18 may generate the element associated with the vector 1818 with a confidence score of 71 %.

The model of FIG. 18 may be continuous bag of words (CBOW) model. The CBOW model may be trained based on a training data set comprising a plurality of input vectors and corresponding output vectors. As the training data
25 set may not be labeled, the training of the CBOW model may be referred to as self-supervised. Before training of the CBOW model, the CBOW model may be initialized with random values assigned to the weights of the neurons. During the training of the CBOW model, the input vectors may be passed through the initialized embedding layer and the output layer and a loss may be determined by comparing the output vector obtained by passing the input vector 1806 through the model to the output vector corresponding to the input vector 1806 as specified by the
30 training data set. Based on the determined loss, backpropagation may be applied to determine the gradients associated with the neurons of the embedding layer 1802 and the output layer 1804 to lower the loss. According to the determined gradients, the weights of the neurons may be updated by using a gradient descent algorithm. If a predetermined loss may be achieved by the CBOW model, the training may be terminated and a trained CBOW model may be obtained. From the trained CBOW model, the embedding layer 1802 may be suitable for embedding
35 input data comprising one or more elements. This embedding layer 1802 may be used in other machine-learning

architectures requiring an embedding layer 1802 such as a transformer encoder, transformer decoder or transformer encoder decoder architecture as described within the context of FIG. 19A, FIG. 19B and FIG. 19C. For training these architectures, a trained embedding layer 1802 may be required. Hence, a model such as a CBOW model may be trained prior to training the transformer encoder, transformer decoder or transformer encoder decoder architecture.

FIG. 19A illustrates an embodiment of a transformer encoder architecture. The transformer encoder comprises an encoder input 1978, one or more encoder blocks 1974, 1914 and an encoder output. The transformer encoder architecture may be derived from the transformer encoder-decoder architecture as known in the art and shown in FIG. 19C. In particular, the transformer encoder may be referred to as X-former. The transformer encoder architecture may correspond to the encoder architecture associated with the transformer encoder-decoder architecture with an additional encoder output instead of connecting the encoder block directly to the decoder of the transformer encoder-decoder architecture. A plurality of transformer encoder architectures are available in the art such as the bi-directional encoder representations from transformers (BERT).

The input data may be received at the encoder input 1978. The encoder input 1978 may apply an input embedding 1902. Applying the input embedding 1902 may refer to passing the input data through an embedding layer e.g. as described within the context of FIG. 18. Further, the encoder input 1978 may apply positional encoding 1904. Applying positional encoding 1904 may refer to adding a positional factor to the embedded input obtained via input embedding. Preferably, the input data may specify a sequence of elements. The positional factor p_{pos} may be indicative of the position of the elements within the sequence. For example, the positional factor p_{pos} may be obtained based on the following equation:

$$p_{pos}(2i) = \sin\left(\frac{pos}{10000^{\frac{2i}{d}}}\right)$$

$$p_{pos}(2i + 1) = \cos\left(\frac{pos}{10000^{\frac{2i}{d}}}\right)$$

where pos may refer to the position of the element within the sequence, i may refer to the dimension associated with the input embedding and d may refer to the dimension of the model, e.g. transformer decoder, transformer encoder or transformer encoder-decoder. This may be referred to as absolute positional embeddings.

Alternatively, the positional encoding may be based on rotary positional embeddings (RoPE). Positional encoding is beneficial since it enables the processing of sequential data without requiring further dimensions indicating the

position of each element. Followingly, the positional encoding 1904 reduces the computational resources needed for embedding the input data. By passing the input data through the encoder input, the input data may be transformed into a second-rank tensor representing the sequence of elements. This second-rank tensor may be referred to as embedded input data. The embedded input data may be processed by the encoder block. The

5 embedded input data may be provided to the layer normalization 1908 by a residual connection. Multi-head self-attention 1906 may be applied to the embedded input data. Multi-head self-attention 1906 may comprise the two components multi-head and self-attention. Self-attention may be understood as being a filter applied to the embedded input data. By applying the filter to the embedded input data, the elements associated with the embedded input data contributing to the to be generated output data may be identified for generating the output

10 data. Hence, the filter may represent the degree of contributing to the to be generated output data by the elements associated with the embedded input data. Applying the filter may be referred to as weighting the elements associated with the embedded input data. This is advantageous specifically regarding long sequences of elements. The filter may be learned and improved during the training by learning to identify the contribution of elements associated with the embedded input data. For example, in the partial sentence "I went to the bakery to

15 buy a" the last word may be generated by the data-driven model such as the transformer encoder. The self-attention may focus the transformer encoder to attend to the word "bakery" and "buy" mostly to generate the word "bread". Self-attention may refer to attention generated based on the input data. Hence, the filter may be determined based on the input data, preferably the embedded input data. The embedded input data may serve as query Q, key K and value V with respect to the self-attention operation. The self-attention may refer to attention

20 based on the received input data. Hence, the filter may be calculated based on the following formula by inserting the respective tensors based on the embedded input data:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

25 where d_k corresponds to the dimension of the key.

For improving the efficiency of the transformer encoder further, the multiple heads are used to apply the filter resulting in the multi-head self-attention 1906. Multi-head self-attention 1906 may comprise applying the filter to two or more parts of the embedded input data. Hence, the tensor may be split into two or more parts and the filter may be applied to the two or more parts separately by two or more heads according to the

30 following equation: $head\ i = Attention(QW_i^Q, KW_i^K, VW_i^V)$

with parameter matrices $W_i^Q \in \mathbb{R}^{d \times d_Q}$, $W_i^K \in \mathbb{R}^{d \times d_K}$, $W_i^V \in \mathbb{R}^{d \times d_V}$ where i may refer to the number of heads, d_V , d_K and d_Q may refer to the dimensions of the value, key and query.

The result of the two or more head may be concatenated according to the following

$$\text{equation: } \text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) W^0$$

where $W_0 \in \mathbb{R}^{hd_V \times d}$ and h may refer to the number of heads.

- 5 The embedded input data may be transformed via the multi-head self-attention 1906 into a context tensor. The context tensor may represent the sequence of elements and the relation between two or more elements of the input data. The context tensor may be a second rank tensor and/or may comprise one or more first rank tensor(s). After the multi-head self-attention 1906 layer normalization 1908 may be applied based on the context tensor and/or the embedded input data from the residual connection. Applying layer normalization 1908 may refer to
- 10 normalizing the context tensor. Normalizing the context tensor may lower the values of the entries of the context tensor. This reduces the computational cost associated with processing the context tensor. Further, it improves the training by contributing the loss to converge and preventing instabilities.
- Layer normalization 1908 may be followed by passing the context tensor to a feed-forward layer 1910 again followed by layer normalization 1912 based on the residual connection to the context tensor and/or the output of
- 15 the feed-forward layer 1910. The feed-forward layer 1910 may be a feed-forward neural network. The feed-forward neural network may comprise of a plurality of fully connected neurons. Passing the context tensor through the feed-forward neural network may result in transforming the context tensor linearly. Additionally or alternatively, the neural network may comprise one or more activation functions such as a rectified linear unit (ReLU). Hence, the neural network may be configured to perform one or more non-linear operations to the context tensor and/or
- 20 transforming the context tensor non-linearly. After the context tensor has been transformed and/or normalized by the feed-forward layer 1910 and the layer normalization 1912, the context tensor may be provided to one or more further encoder blocks 1914. Having passed the context tensor through the feed-forward layer 1910 may adapt the context tensor for the processing by a further attention layer of the one or more further encoder blocks 1914 for applying a self-attention filter, preferably multi-head self-attention 1906. The context vector after being
- 25 transformed by the layer normalization 1912 and the feed-forward layer 1910 may be referred to as hidden state. The encoder output 1976 comprises a linear layer 1916 and a softmax layer 1918. The linear layer 1916 may transform the context vector into a logits vector. The linear layer may be fully-connected. The logits vector obtained by passing the context tensor through the linear layer 1916 may be passed through the softmax layer 1918. Passing the logits vector through the softmax layer 1918 may refer to applying the softmax function to the
- 30 logits vector. Applying the softmax function to the logits vector may result in a probability distribution of one or more elements corresponding to the sequence of elements in the input data. The probability distribution of the one or more element(s) may be confidence score(s) associated with the one or more element(s). From the probability distribution based on predefined selection criteria, one or more elements may be chosen. The one or more chosen elements may be referred to as the one or more elements generated by the transformer encoder. The one or more
- 35 generated elements may be provided to the encoder input for generating further one or more elements

corresponding to the sequence of the input data and the one or more elements generated by the transformer encoder as described within the context of FIG. 20.

FIG. 19B illustrates an embodiment of a transformer decoder architecture.

5

The transformer decoder comprises a decoder input 1984, one or more decoder blocks 1980, 1932 and a decoder output 1992. The transformer decoder architecture may be derived from the transformer encoder-decoder architecture as known in the art and shown in FIG. 19C. The transformer decoder may be referred to as X-former. The transformer decoder architecture may correspond to the decoder architecture associated with the transformer encoder-decoder architecture independent of receiving one or more hidden states from the encoder of the transformer encoder-decoder. A plurality of transformer decoder architectures are available in the art such as the generative pretrained transformers (GPT).

10

The decoder input 1984 may apply input embedding 1920 and positional encoding 1922 analogous to analogous to the input embedding 1902 and the positional encoding 1904 as described within the context of FIG. 19A.

15

The decoder block 1980 may comprise the layer normalizations 1926, the masked multi-head self-attention 1924, the feed-forward layers 1928 and/or the layer normalization 1930. The embedded input data resulting from passing the input data through the decoder input 1984 may be provided to the layer normalization 1926 via a residual connection. Further, masked multi-head self-attention 1924 may be applied to the embedded input data. Masked multi-head self-attention 1924 corresponds to the multi-head self-attention 1906 as described within the context of FIG. 19A with additionally masking a part of the embedded input data associated with elements later in the sequence than the element to be generated. Additionally or alternatively, the part of the input data associated with elements later in the sequence than the element to be generated may not be received and/or transformed into the embedded input data. Thus, the transformer decoder may be suitable for generating a subsequent element to a sequence, whereas the transformer encoder may be suitable for generating a missing element in within one sequence and/or between two or more sequences. Therefore, the transformer encoder may be configured to perform classification tasks. The transformer decoder may be configured to generate text.

20

25

Similar to the transformer encoder as described within the context of FIG. 19A, a context tensor may be generated by applying the masked multi-head self-attention 1924 and the layer normalization 1926. The context tensor may be provided to the layer normalization 1930 via a residual connection. Further, the feed-forward layer 1928 and the layer normalization 1930 may be analogous to the feed-forward layer 1910 and the layer normalization 1912 as described within the context of FIG. 19A. The context tensor may be provided to one or more further decoder blocks 1932.

30

35

The decoder output 1992 may comprise of a linear layer 1934 and a softmax layer 1936. The linear layer 1934 and the softmax layer 1936 may be analogous to the linear layer 1916 and the softmax layer 1918 as described within the context of FIG. 19A.

FIG. 19C illustrates an embodiment of a transformer encoder-decoder architecture. The transformer encoder-decoder may comprise the encoder input 1988, the one or more encoder blocks 1986, 1964, the decoder input 1994, the decoder block 1990 and the decoder output 1992. The encoder input 1988 may correspond to the encoder input 1978 of FIG. 19A. The one or more encoder block 1986, 1964 may correspond to the one or more encoder blocks 1974, 1914 of FIG. 19A. The decoder input 1994 may correspond to the decoder input 1984 of FIG. 19B.

The decoder block 1990 may comprise a masked multi-head self-attention 1970, a layer normalization 1972, a feed-forward layer 1938 and a layer normalization 1940 analogous to the masked multi-head self-attention 1924, the layer normalization 1926, the feed-forward layer 1928 and the layer normalization 1930 as described within the context of FIG. 19B. The decoder block 1990 may further comprise a multi-head self-attention 1950 and a layer normalization 1948. Analogous to the description of FIG. 19B, the context tensor may be obtained from the masked multi-head self-attention 1970 and the layer normalization 1972. Multi-head self-attention 1950 analogous to the multi-head self-attention 1906 of FIG. 19A may be applied to the context vector obtained from the layer normalization 1972 and the hidden states of the one or more encoder blocks 1986, 1964. Layer normalization 1948 may be applied to the context vector obtained from the multi-head self-attention 1950 and the context vector obtained from the layer normalization 1972 provided via a residual connection. The context vector resulting from the layer normalization 1948 may be processed via the feed-forward layer 1938 and the layer normalization 1940 analogous to the description of FIG. 19B. The context vector resulting from the layer normalization 1940 may be provided to further decoder blocks 1942 analogous to the decoder block 1990. The context vector obtained from the one or more decoder blocks 1990, 1942 may be provided to the decoder output 1992. The decoder output 1992 may correspond to the decoder output 1982 of FIG. 19B.

With the above-described architecture, the transformer encoder-decoder may receive and process input data at the encoder input 1988 and the one or more encoder blocks 1986, 1964 and the decoder block 1990 and the decoder output 1992. Based on the input data, the transformer encoder-decoder may generate output data part by part or sequentially. The sequentially generated output data may be provided to and/or may be processed by the decoder input 1994, the one or more decoder blocks 1990, 1942 and the decoder output 1992. Preferably, a sequence may be provided to the encoder input 1988 and after having generated at least a part of the output data, the decoder input 1994 may be provided with at least the part of the elements of the output data already generated. By doing so, the next elements of the output data may be generated with a higher accuracy by taking the input data and the generated output data into account since more data is received by the transformer encoder-decoder may be received over time.

Because of the transformer encoder-decoder architecture, the transformer encoder-decoder may be configured to transform a sequence into another representation of the sequence. An example for transforming one sequence into another representation may be translation of one sentence into another language. A plurality of transformer encoder-decoders are available in the art such as BART, T5 or the like.

5

In an embodiment, the layer normalization 1908, 1912 may be applied prior to the masked multi-head self-attention 1924, multi-head self-attention 1906 and/or the feed-forward layer 1910 in the transformer decoder, the transformer encoder and/or the transformer encoder-decoder. By doing so, the computational resources for applying the multi-head self-attention 1906 and/or the feed-forward layer 1910 to the embedded input data and/or

10

the context tensor may be decreased as the entries of the respective tensors may be lower after normalization. In an embodiment, the decoder output 1992 may comprise of a classification neural network, further feedforward layers, convolutional layers, fully connected layers or the like. For example, the transformer encoder-decoder may be configured to choose between a plurality of options. For this purpose, the transformer encoder-decoder may be provided with three different input data sets and may classify the context vectors obtained from the one or more

15

decoder blocks 1990 via one or more linear layers. Followingly, the architecture may be extended depending on the use case to be solved.[1]

FIG. 20 illustrates an embodiment of training and/or deploying the transformer encoder, the transformer decoder and/or the transformer encoder-decoder.

20

The encoder/decoder/encoder-decoder architecture 2002 may correspond to the transformer decoder, the transformer encoder and/or the transformer encoder-decoder as describe within the context of FIG. 19A- FIG. 19C.

The output data generated by the encoder/decoder/encoder-decoder architecture 2002 may comprise of one or

25

more elements, in particular a sequence of elements. The previously generated elements of the output data may be provided as input for generating the next element in the sequence of the output data. In the example of FIG. 20, the input data may comprise of N elements, in particular input tokens. An input token may be a token dedicated to be inputted into a data-driven model such as the transformer decoder, the transformer encoder or the transformer encoder-decoder. The output data to be generated may comprise of M

30

elements. The encoder/decoder/encoder-decoder architecture 2002 may generate one element of the output data based on receiving the input data and optionally previously generated elements of the output data at a timestep. Hence, for generating M elements M time steps are required. A time step comprises of providing input 2010, 2012, 2014 to the encoder/decoder/encoder-decoder architecture 2002 and receiving output data 2004, 2008, 2006 from the encoder/decoder/encoder-decoder architecture 2002. In a first timestep, the input 2010 may comprise of N

35

input tokens. The N input tokens may be associated e.g. with N words, stems or endings. Preferably, the N input

tokens may specify a question. One or more input tokens may specify the beginning of the sequence of tokens and/or the end of the sequence of tokens. The input 2010 may be processed by the encoder/decoder/encoder-decoder architecture 2002. Based on the input 2010 at least a part of the output data 2004 may be generated. The at least a part of the output data may comprise a first output token. In the next timestep, the generated first output
5 token may be provided together with the input 2012. Specifically, where the input 2012 may be received by a transformer encoder-decoder the input tokens may be received at the encoder input 1988 and the first output token may be received at the decoder input 1994. Where the input 2012 may be received by the transformer encoder, the input 2012 may be received by the encoder input 1978 and analogously regarding the transformer decoder and the decoder input 1984. Based on the input 2012, the output data 2008 comprising the first output
10 token and a second output token may be generated. Generating the output data 2008 based on the input 2012 may refer to generating the second token based on the first token and the N input tokens, wherein the first token may have been generated based on the N input tokens. This process may be repeated until the last token in the sequence of the output data 2006 may be generated. Preferably, the last token may be an end token. The end token may terminate the generation of a further output token.

15 Similarly, to the data processing during deployment of the encoder/decoder/encoder-decoder architecture 2002, the encoder/decoder/encoder-decoder architecture 2002 may be trained. The training data set may comprise a plurality of sequences comprising a plurality of elements. The sequences may be associated with the input data and/or the output data. Additionally or alternatively, the sequences may be independent of the input data and/or the output data. For example, where the input data and the output data may refer to chemical compositions
20 represented via text, the training data set may comprise sequential text data independent of chemical compositions. In this example, the training data set may comprise sequences of words originating from a conversation. In an embodiment, the training data set may comprise at least partially input data sets and/or output data sets.

The training may be initialized by initializing the encoder/decoder/encoder-decoder architecture 2002. In an
25 embodiment, the parameters associated with the encoder/decoder/encoder-decoder architecture 2002 may be initialized randomly. Additionally or alternatively, the input embedding of the encoder/decoder/encoder-decoder architecture 2002 may be obtained by training a CBOW model or a skip gram model as described within the context of FIG. 18. The trained embedding layer may be used during training. The parameters associated with the embedding layer may be kept constant and/or may be updated after a predefined number of training epochs. By
30 doing so, the number of parameters to be updated is lower enabling a faster and less computational resources-consuming training. Further, the accuracy associated with the embedding layer may be constant and/or may be increased by avoiding error compensation in relation to the just initialized encoder/decoder/encoder-decoder architecture 2002.

During the training of the encoder/decoder/encoder-decoder architecture 2002, at least a part of the sequences of
35 the training data set may be provided to the encoder/decoder/encoder-decoder architecture 2002 one by another

and one or more elements may be generated based on the sequences of the training data set one by another. The elements generated based on the sequences may follow the elements of the parts of sequences the encoder/decoder/encoder-decoder architecture 2002 may have been provided with. The generated one or more elements may be compared to the one or more elements following the at least a part of the sequences provided to the encoder/decoder/encoder-decoder architecture 2002 as specified by the training data set. Hence, during the training the encoder/decoder/encoder-decoder architecture 2002 may generate a guess on the next element and the guess on the next element in a sequence may be compared to the ground truth specifying the actual next element according to the training data set. Based on the guess on the next element and the ground truth a loss may be determined. The loss may define the similarity between the guess on the next element and the ground truth. The loss may be determined by forming a vector dot product between the token associated with the one or more elements and the token associated with the ground truth. A loss unequal to zero may result in updating the parameters associated with encoder/decoder/encoder-decoder architecture 2002. Preferably the parameters associated with the encoder/decoder/encoder-decoder architecture 2002 may be independent of the embedding layer. For example, the parameters associated with the encoder/decoder/encoder-decoder architecture 2002 may be weights of the neurons of the encoder/decoder/encoder-decoder architecture 2002.

Based on the determined loss, backpropagation may be applied to determine the gradients associated with the parameters of the parameters associated with encoder/decoder/encoder-decoder architecture 2002 to lower the loss. According to the determined gradients, the parameters associated with the encoder/decoder/encoder-decoder architecture 2002, preferably the weights of the neurons associated with the encoder/decoder/encoder-decoder architecture 2002, may be updated by using a gradient descent algorithm.

The training data set may be unlabeled. The sequences of elements within the training data set may inherently comprise the ground truth for determining the loss with respect to the one or more elements generated during the training of the encoder/decoder/encoder-decoder architecture 2002. Hence, the encoder/decoder/encoder-decoder architecture 2002 may be trained self-supervised. This is advantageous since time and resources for creating a labeled training data set may be saved. Furthermore, this enables the usage of large training data sets associated with a size of several tera bytes. Consequently, the data-driven model may be accurate in generating elements of a sequence. In addition, the large training data set enables few shot predictions or even zero shot predictions. Hence, the data-driven model(s) trained as described above are versatile contributing to saving resources needed for training and/or hosting a plurality of purpose-driven models such as convolutional neural networks. The training described above may be referred to as pretraining. The data-driven model may be configured to perform few shot or even zero shot predictions with respect to a plurality of use cases after pretraining. The performance of the data-driven model may be increased further by additional training referred to as finetuning.

FIG. 21 illustrates an embodiment of input embedding. Where the sequence of elements associated with the input data, preferably comprised in the input data, may be of one type, the input embedding 1902, 1920, 1952, 1966 as described within the context of FIG. 19A - 2C may be used. For example, a type of input data may be text where the elements may be associated with at least a part of a word, a punctuation character, a start token specifying the beginning of one or more sequences associated with the input data and/or the end token. In another example, the input data may be at least partially numerical. Hence, the input data may comprise a plurality of numbers. Numerical input data may be for example tabular data. Tabular data may specify one or more rows and/or one or more columns. Hence, the tabular data may comprise one or more cells, wherein the cells may be associated with one or more numerical values.

Numerical input data may require a different embedding than text input data. Input embeddings for numerical input data may comprise a token embedding, a positional embedding, a column embedding, a row embedding or a combination thereof.

Applying a token embedding to one or more elements, in particular tokens associated with the input data may result in a machine-processable representation associated with the one or more elements, in particular tokens.

Applying the token embedding to one or more elements may refer to passing the one or more elements through the embedding layer, e.g. as described within the context of FIG. 18. Hence, token embeddings may specify the one or more elements, in particular tokens in a machine-processable representation. For example, the token embedding may transform a numerical value into a vector. This is advantageous since this representation can be enriched by further information such as the position of the token within the sequence and/or within a table associated with the sequence of tokens. The positional embedding may be analogous to the positional embedding as described within the context of FIG. 18, FIG. 19A-2C. Where the input data may be tabular data, column embedding may be applied. Applying a column embedding to one or more elements, in particular tokens associated with the input data may result in a machine-processable representation specifying the location of the one or more elements within a table 2102, preferably within the columns of the table 2102. Applying the column embedding may refer to adding a column factor to the input data embedded via token embeddings, in particular the embedded input data. The column factor may be the same for elements associated with the same column and/or may differ between two or more elements associated with different columns. Analogous, row embeddings may be applied where the input data may be tabular data. Applying a row embedding to one or more elements, in particular tokens associated with the input data may result in a machine-processable representation specifying the location of the one or more elements within a table 2102, preferably within the rows of the table 2102. Applying the row embedding may refer to adding a column factor to the input data embedded via token embeddings, in particular the embedded input data. The row factor may be the same for elements associated with the same row and/or may differ between two or more elements associated with different rows.

In an embodiment, input data may be at least partially numerical and at least partially text. Hence, the input data may comprise two or more types of data. A type of data may refer to a modality. Followingly, different embeddings may be applied to the input data. To parts of the input data comprising text the input embedding referred to in FIG. 18, FIG. 19A-2C may be applied. To parts of the input data being numerical token embeddings, positional embeddings, column embeddings and row embeddings may be applied. Further, segment embeddings may be applied to the input data independent of the type of input data. The segment embedding may specify the type of input data one or more elements may be associated to. For example, if the input data comprises of text and numbers, the input data may comprise of two types of input data. Applying the segment embedding to the input data may refer to adding a segment factor to the input data, preferably the embedded input data and/or the input data after having applied the token embedding. The segment factor may specify the type of data associated with the one or more elements. The segment factor may be the same for one or more elements associated with the same type of input data and/or may differ between two or more elements associated with different types of input data.

Applying the token embedding, the positional embedding, the segment embedding, the column embedding, the row embedding or a combination thereof may result in embedded input data and/or may be the output of any one of the encoder input 1978, 1984, 1988 or decoder input 1984, 1994. The data obtained by applying the token embedding, the positional embedding, the segment embedding, the column embedding, the row embedding or a combination thereof may be processed by the encoder block 1974, 1986, decoder block 1980, 1990, encoder output 1976, decoder output 1992, 1982.

FIG. 22 illustrates an embodiment of input embedding.

Input data to the data-driven model, in particular to the encoder input and/or the decoder input as described in the context of FIG. 19A-C, may comprise image data. The data-driven model may be parametrized to receive image data. For processing image data as input data, the data-driven model may comprise one or more encoder blocks and/or one or more decoder blocks and/or one or more encoder outputs and/or one or more decoder outputs as described within the context of FIG. 19A-C. FIG. 22 may show an embodiment of an encoder input and/or a decoder input. When processing image data, the encoder input and/or the decoder input of the data-driven model may be as described within the context of FIG. 22. The encoder input and/or decoder input may comprise one or more linear projection layers 2214 for a linear projection of one or more images, preferably one or more partial images, more preferably a sequence of two or more partial images. The one or more linear projection layers 2214 may be suitable for changing the dimension of the one or more received images, preferably one or more partial images, preferably passing the one or more images, preferably partial images, through the one or more linear

projection layers 2214 may result in applying image embedding, preferably partial image embedding to the one or more images and/or partial images.

Furthermore, when a sequence of two or more images and/or partial images may be received, positional embedding may be applied to the sequence, preferably by passing the sequence of one or more images and/or partial images through the one or more linear projection layers 2214. Applying positional embedding may refer to adding a positional factor. The positional factor may be different depending on the position of the image and/or the partial image within the sequence. In particular, the positional factor added to a first element of the sequence may be different to the positional factor added to a second element of the sequence. The first element of the sequence may be a first image and/or first partial image. The second element of the sequence may be a second image and/or a second partial image.

The representation of the one or more images, preferably one or more partial images, may be obtained based on the following equation:

$$z_0 = \left[x_{class}; x_p^1 E; x_p^2 E, \dots, x_p^N E \right] + E_{pos}, \quad E \in \mathbb{R}^{(P^2 C) \times D}, \quad E_{pos} \in \mathbb{R}^{(N+1) \times D}$$

where x_{class} is the image class embedding 2228, x_p^N is the n-th image, in particular partial image in the

sequence, z_0 is the representation of the one or more images, preferably one or more partial images, (H,W) are the resolution of the image, in particular the image the partial images are generated on, C is the number of channels associated with the one or more image, in particular the one or more partial images and D is the dimension of the representation of the one or more images, preferably one or more partial images. Applying the partial image embedding may refer to forming the product of x_p^N with E above-described equation. Applying the

positional embedding may refer to adding the factor E_{pos} according to the above-described equation.

By doing so, text-based data, numerical data, tabular data, image data or the like may be processed by one data-driven model.

The present disclosure has been described in conjunction with preferred embodiments and examples as well.

However, other variations can be understood and effected by those persons skilled in the art and practicing the claimed subject-matter, from the studies of the drawings, this disclosure and the claims. Notably, in particular, the any steps presented can be performed in any order, i.e. the present disclosure is not limited to a specific order of these steps. Moreover, it is also not required that the different steps are performed at a certain place or at one node of a distributed system, i.e. each of the steps may be performed at different nodes using different equipment/data processing.

As used herein „determining“ also includes „initiating or causing to determine“, „generating“ also includes „initiating and/or causing to generate“ and „providing“ also includes „initiating or causing to determine, generate, select, send and/or receive“. „Initiating or causing to perform an action“ includes any processing signal that triggers a computing node or device to perform the respective action.

In the claims as well as in the description the word "comprising" or "including" or similar wording does not exclude other elements or steps and shall not be construed limiting to the elements or steps lined out. The indefinite article "a" or "an" does not exclude a plurality. A single element or other unit may fulfill the functions of several entities or items recited in the claims. The mere fact that certain measures are recited in the mutual different dependent claims does not indicate that a combination of these measures cannot be used in an advantageous implementation or further elements may be included.

Providing in the scope of this disclosure may include any interface configured to provide data. This may include an application programming interface, a human-machine interface such as a display and/or a software module interface. Providing may include communication of data or submission of data to the interface, in particular display to a user or use of the data by the receiving entity.

Any disclosure and embodiments described herein relate to the methods, the systems, devices, the computer program element lined out above and vice versa. Advantageously, the benefits provided by any of the embodiments and examples equally apply to all other embodiments and examples and vice versa.

The present disclosure has been described in conjunction with preferred embodiments and examples as well. However, other variations can be understood and effected by those persons skilled in the art and practicing the claimed invention, from the studies of the drawings, this disclosure and the claims. Notably, in particular, the any steps presented can be performed in any order, i.e. the present invention is not limited to a specific order of these steps. Moreover, it is also not required that the different steps are performed at a certain place or at one node of a distributed system, i.e. each of the steps may be performed at different nodes using different equipment/data processing.

As used herein „determining“ also includes „initiating or causing to determine“, „generating“ also includes „initiating and/or causing to generate“ and „providing“ also includes „initiating or causing to determine, generate, select, send and/or receive“. „Initiating or causing to perform an action“ includes any processing signal that triggers a computing node or device to perform the respective action.

In the claims as well as in the description the word "comprising" does not exclude other elements or steps. The indefinite article "a" or "an" and the definite article "the" does not exclude a plurality. In particular, indefinite article "a" or "an" may be replaced with one or more and the definite article "the" may be replaced with the one or more. A single element or other unit may fulfill the functions of several entities or items recited in the claims. The mere fact that certain measures are recited in the mutual different dependent claims does not indicate that a combination of these measures cannot be used in an advantageous implementation.

Any disclosure and embodiments described herein relate to the methods, the systems, devices, the computer program element lined out above and vice versa. Advantageously, the benefits provided by any of the embodiments and examples equally apply to all other embodiments and examples and vice versa.

CLAIMS

1. A method, in particular computer-implemented method, for generating chemical product data characterizing a chemical product with one or more target properties, the method comprising:
 - 5 providing a request for providing the chemical product with the one or more target properties, wherein the request includes unstructured data and is associated with one or more target properties of the chemical product,
 - providing functional specification data related to one or more functions of one or more operating engine(s) for providing the chemical product with the one or more target properties,
 - 10 providing one or more input data structure related to input data suitable for being provided to the one or more operating engine(s),
 - providing a task instruction including unstructured data related to the request, the one or more input data structure(s) and the functional specification to one or more data-driven model(s), wherein the one or more data-driven model(s) are configured to generate operating input data for one or more
 - 15 selected operating engine(s) in relation to the provided one or more target properties, wherein the operating input data includes structured data for triggering the selected operating engine,
 - providing the operating input data to the at least one selected operating engine for generating the chemical product data, wherein the operating engine is configured to generate at least a part of the chemical product data in response to providing the operating input data,
 - 20 providing the generated chemical product data for producing and/or processing the chemical product with the one or more target properties.
2. The method of claim 1, wherein the indication on the target chemical product and/or the operating input data includes a digital representation of the target chemical product.
- 25 3. The method of claim 1 or 2, wherein the two or more operating engine(s) include at least one operating engine configured to provide the chemical product data and at least one operating engine configured to provide data of another type than the chemical product data.
- 30 4. The method of any one of claims 1 to 3, wherein the functional specification data includes a digital representation of the one or more functions per operating engine and a relation of the one or more functions to the two or more operating engines corresponding to the one or more functions.
- 35 5. The method of any one of claims 1 to 4, wherein providing the selection task instruction includes mapping the selection task instruction to a numerical representation of the selection task instruction and wherein the

one or more selection model(s) are configured to map the numerical representation of the selection task instruction to a numerical representation of the operating input data.

- 5 6. The method of any one of claims 1 to 5, wherein the selection model is further configured to map a numerical representation of the selection task instruction to a context selection task instruction and map the context selection task instruction to a numerical representation of the operating input data, wherein the context selection task instruction is obtained by processing the numerical representation of the selection task instruction by one or more matrix operation(s) associated with the selection model.
- 10 7. The method of any one of claims 1 to 6, further comprising providing a validation task instruction related to the request and the operating input data to a validation model for validating the at least one selected operating engine corresponds to the request, wherein the validation model is configured to follow task instructions, and wherein optionally the validation model may be the selection model.
- 15 8. The method of any one of claims 1 to 7, wherein the request and/or the selection task instruction comprises unstructured data.
- 20 9. The method of any one of claims 1 to 8, wherein the selection model is a pretrained data-driven model, wherein the pretrained data-driven model is configured to perform a plurality of different tasks according to a plurality of different task instructions.
- 25 10. The method of any one of claims 1 to 9, wherein the selection model is a finetuned data-driven model, wherein the finetuned data-driven model is obtained by further training a pretrained data-driven model based on training data comprising selection task instructions and corresponding operating input data, wherein the pretrained data-driven model is configured to perform a plurality of different tasks according to a plurality of different task instructions.
- 30 11. The method of any one of claims 1 to 10, wherein providing the selection task instruction for generating the operating input data comprises
providing the selection task instruction related to the request and the functional specification data to the selection model configured to generate model output data related to the selected operating engine and the one or more target properties, and
providing the generated model output data to a structure model configured to generate operating input data from the model output data.
- 35

12. An apparatus comprising:
a processor; and
a memory storing instructions that, when executed by the processor, configure the apparatus to perform the steps of any one of the methods of any one of claims 1 to 11.

5

13. Use of chemical product data as obtained by any one of claims 1 to 11 for producing and/or processing a chemical product.

10

14. Use of a selection task instruction according to any one of claims 1 to 11 for processing a request for providing chemical product data associated with a target chemical product according to any one of claims 1-11.

15

15. Use of a selection model according to any one of claims 1 to 11 for providing chemical product data for producing and/or processing a target chemical product.

1/27

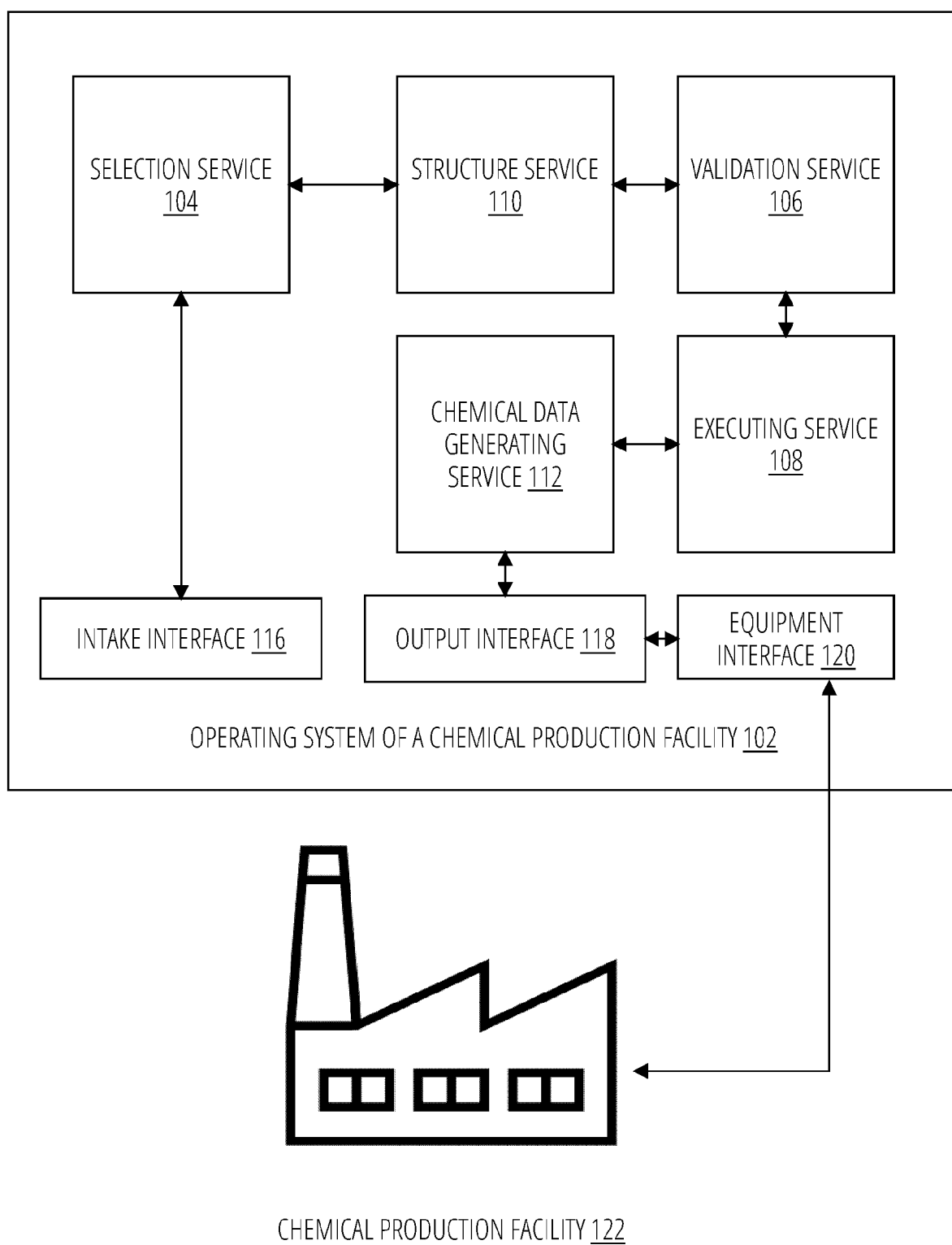


FIG. 1

2/27

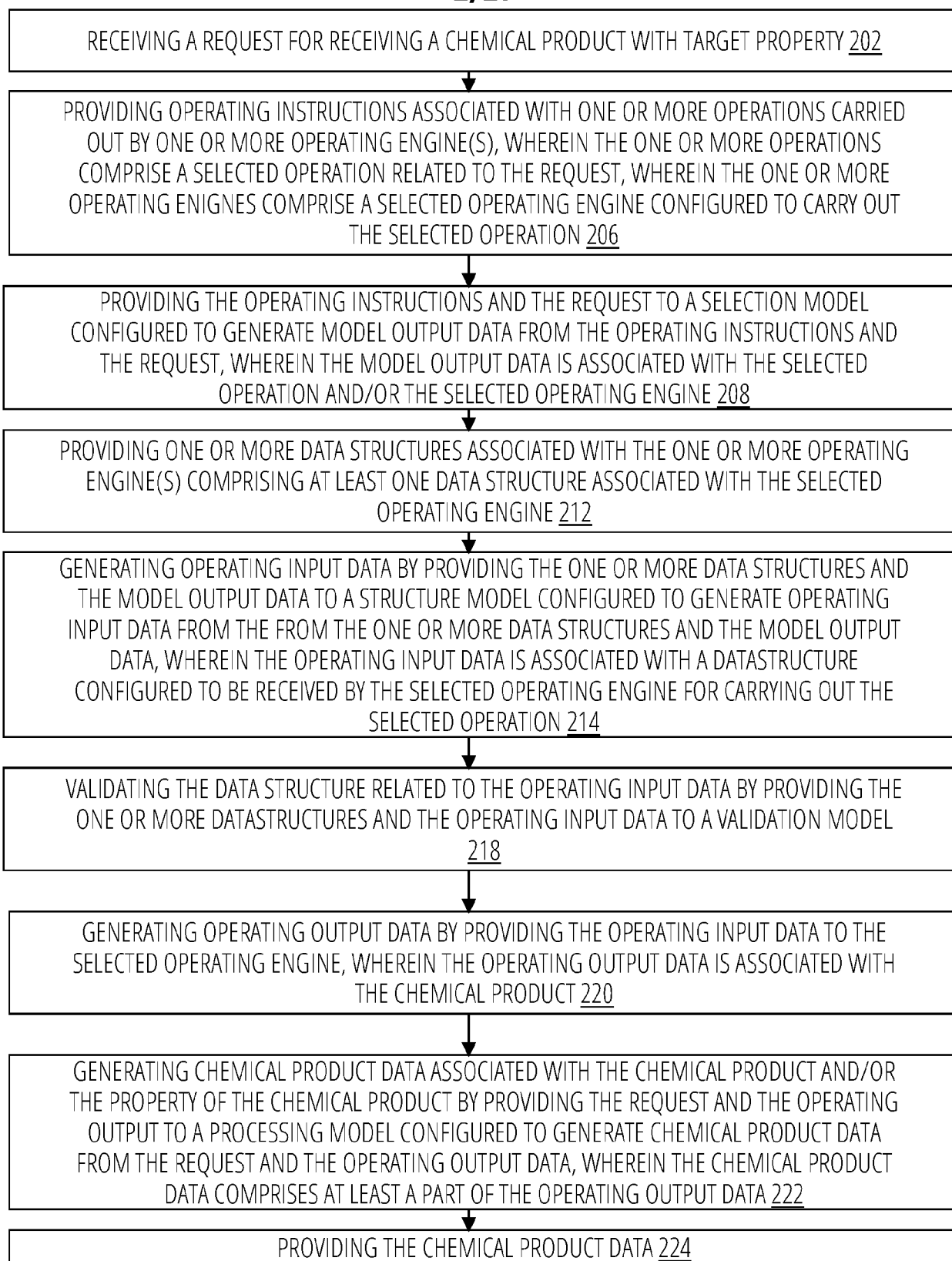
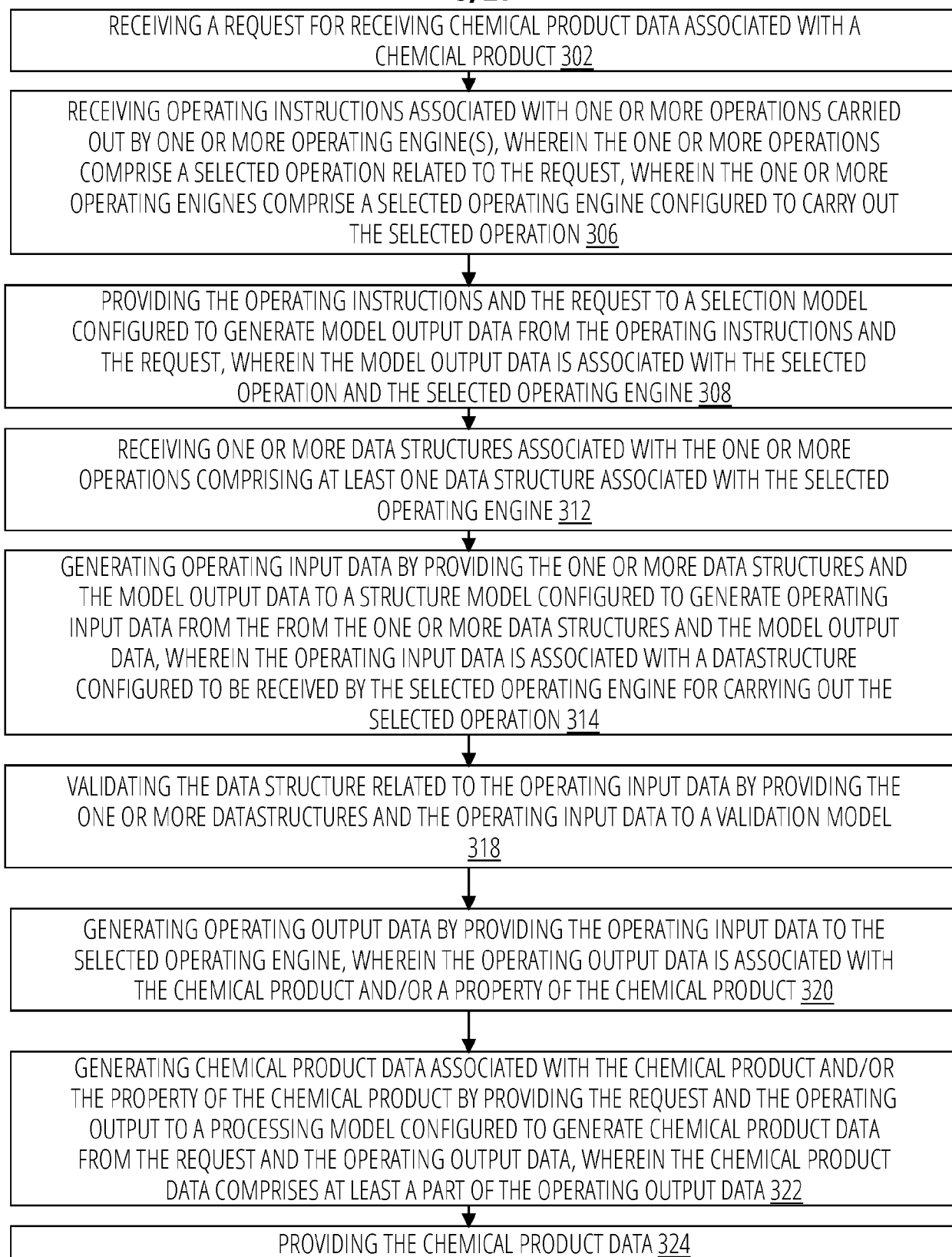


FIG. 2

3/27**FIG. 3**

4/27

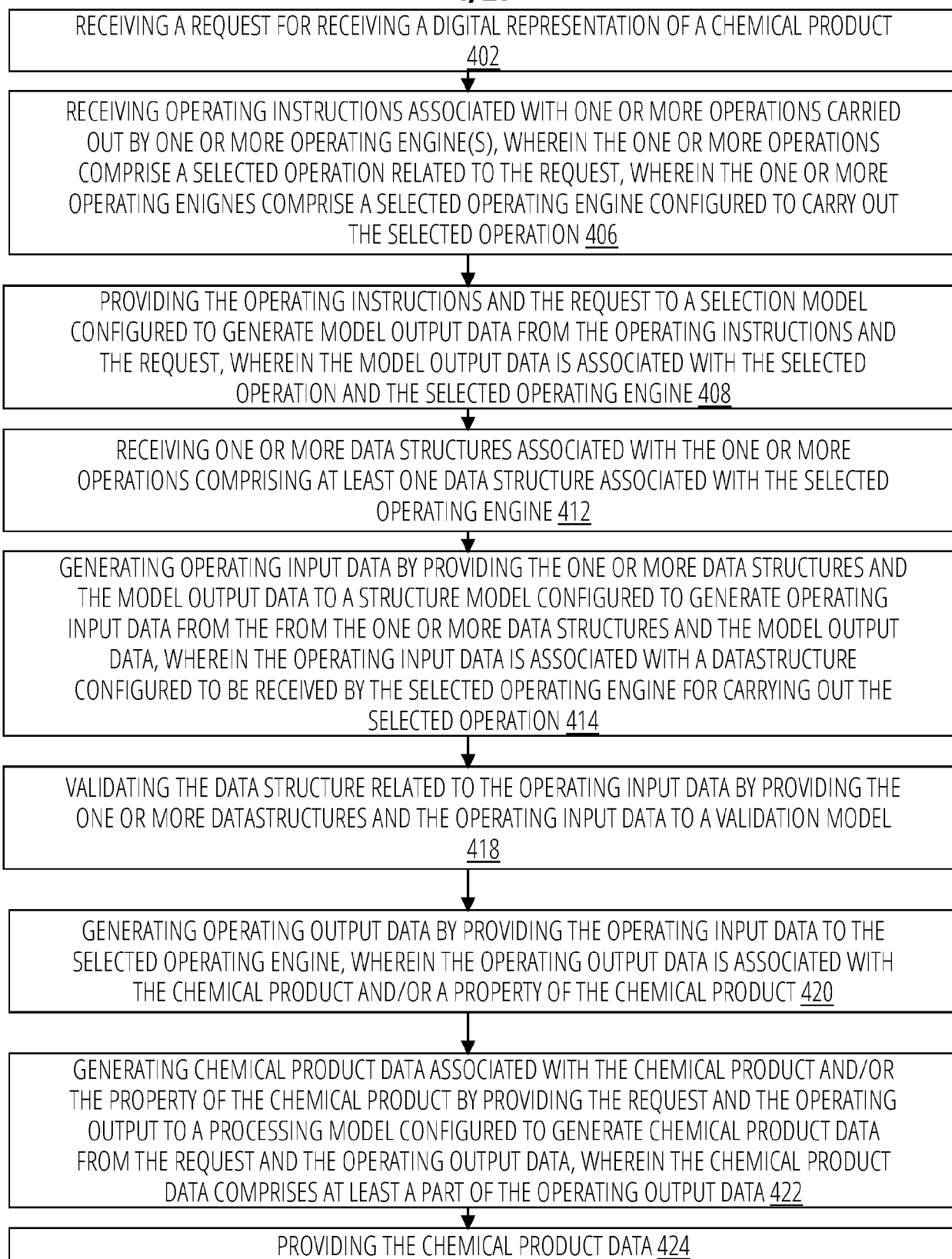


FIG. 4

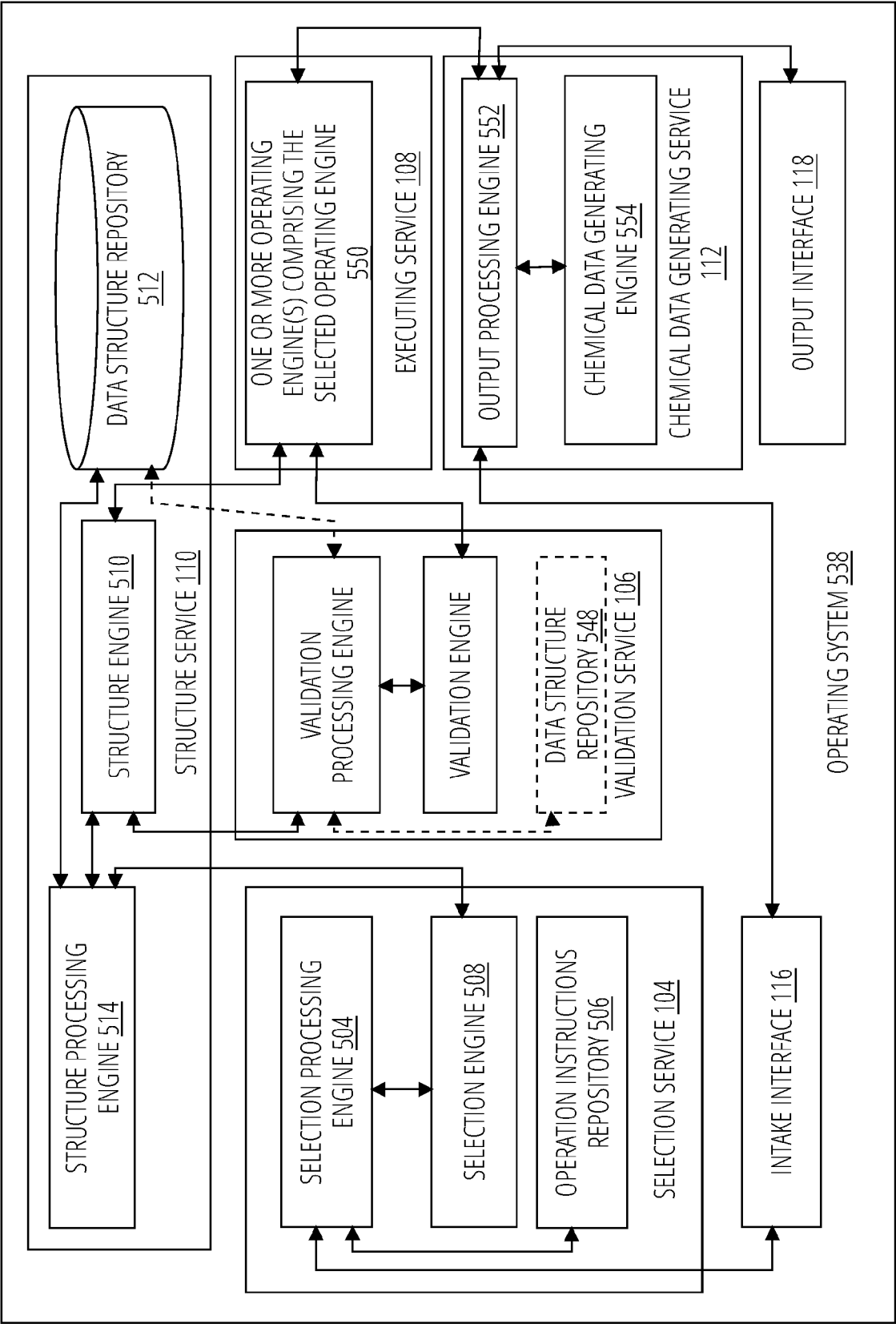


FIG. 5

6/27

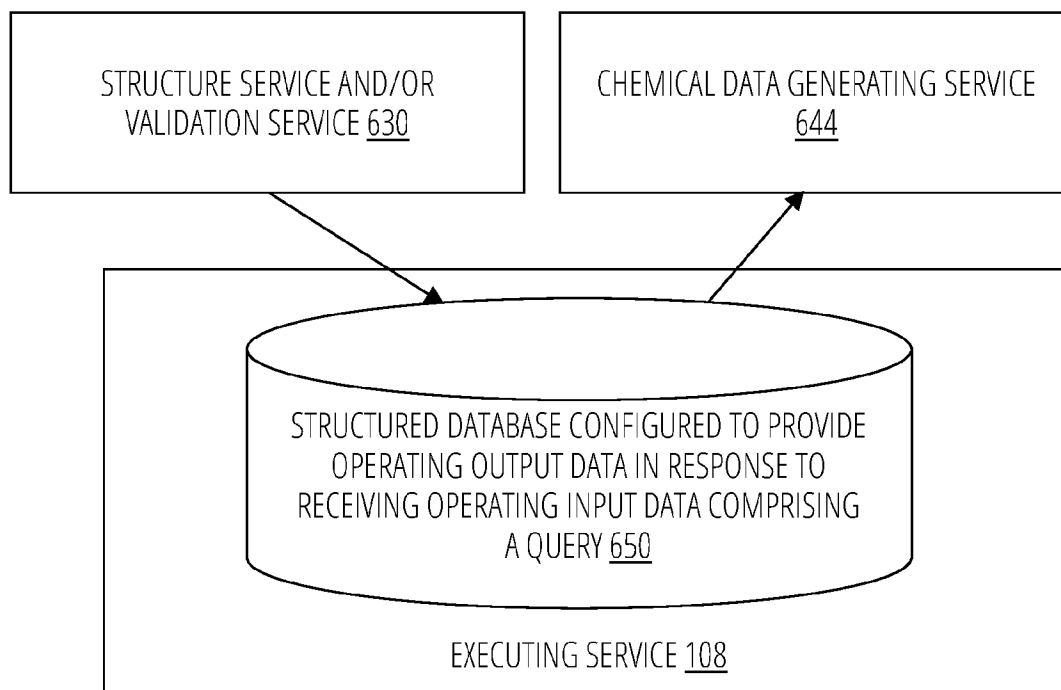


FIG. 6A

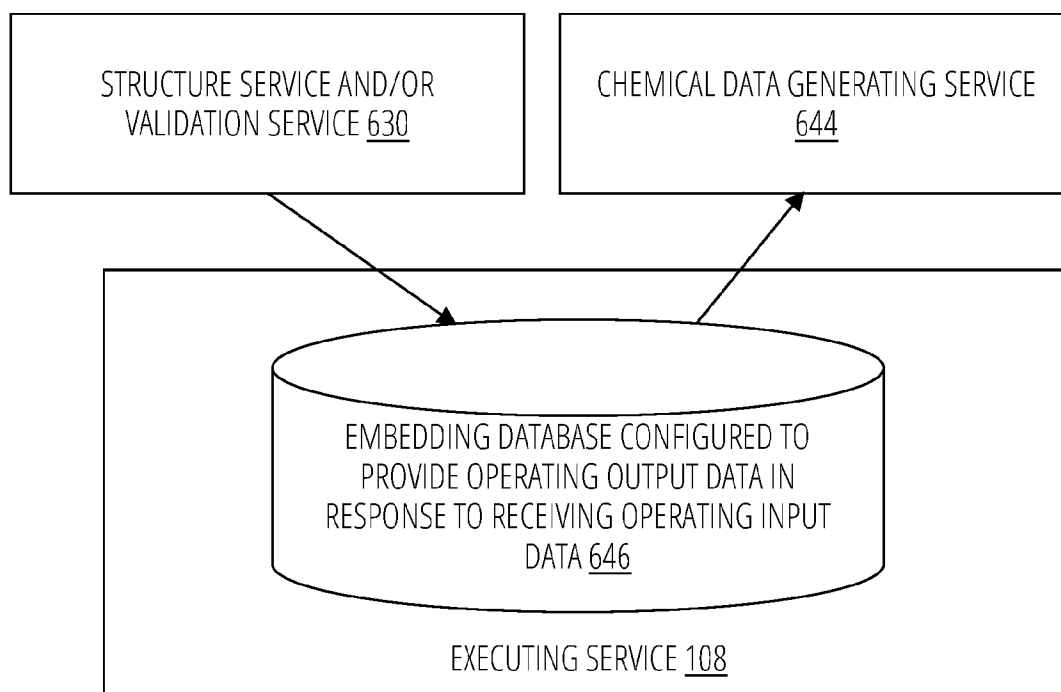


FIG. 6B

7/27

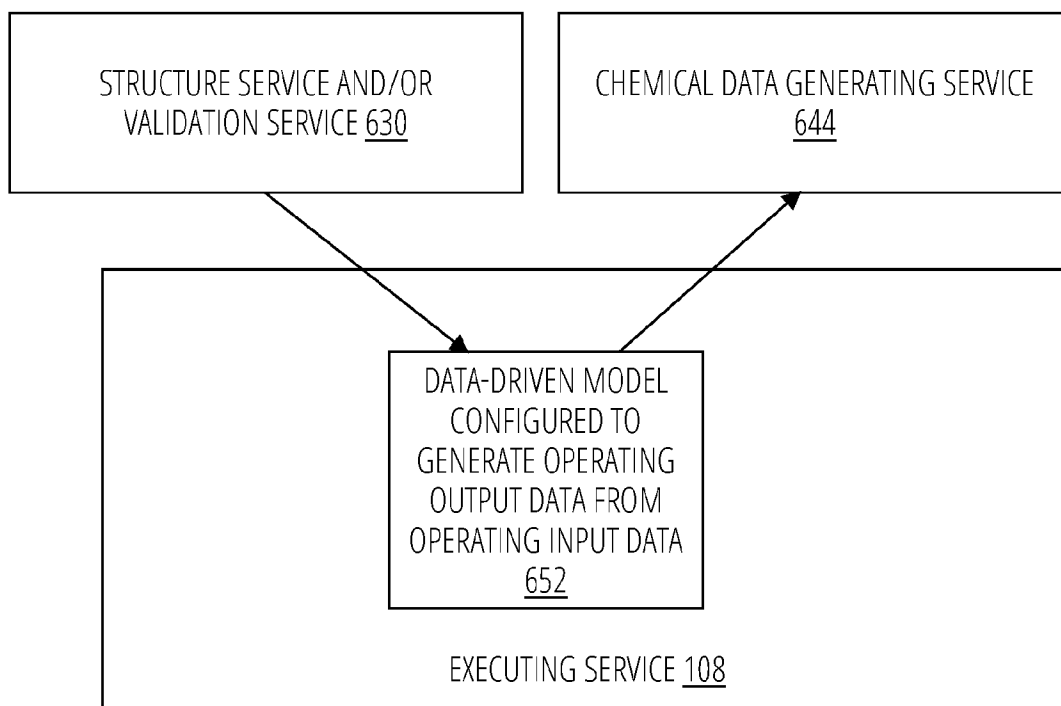


FIG. 6C

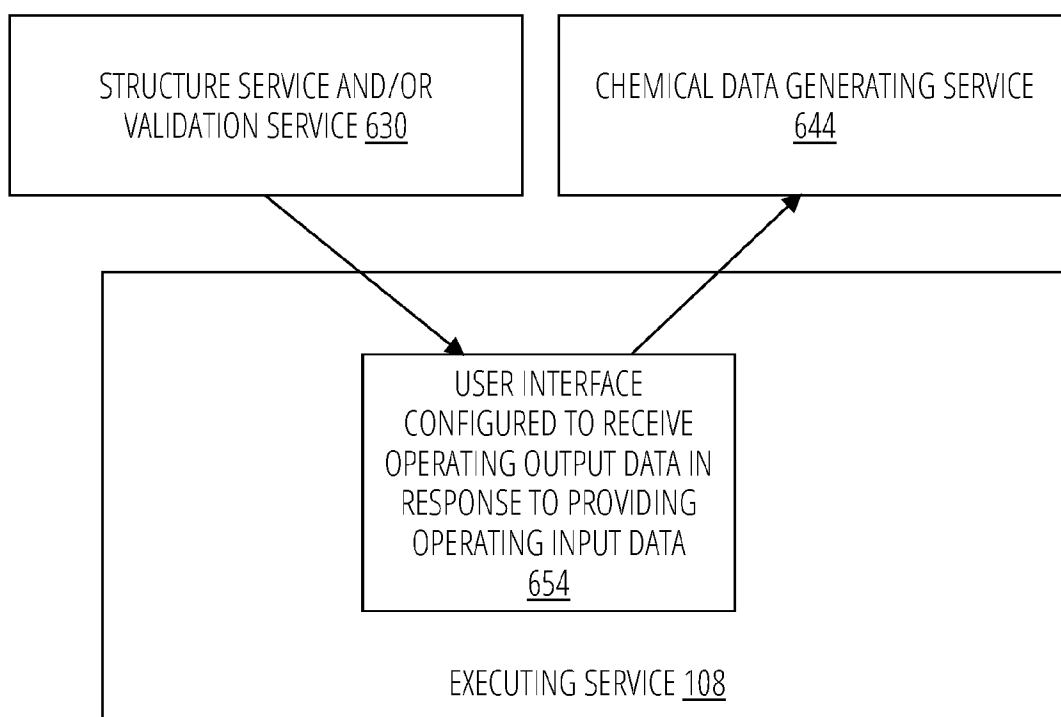


FIG. 6D

8/27

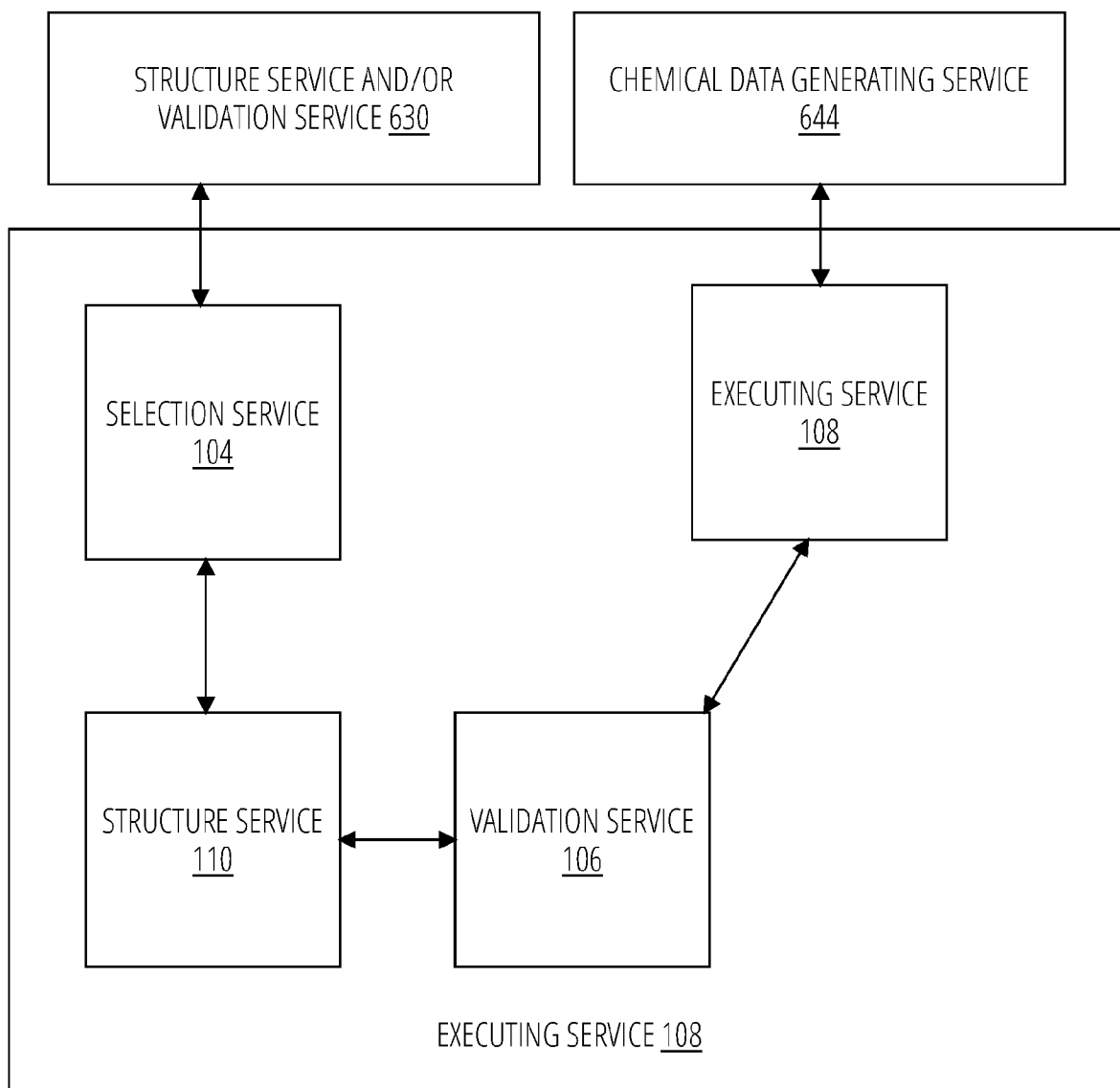


FIG. 6E

9/27

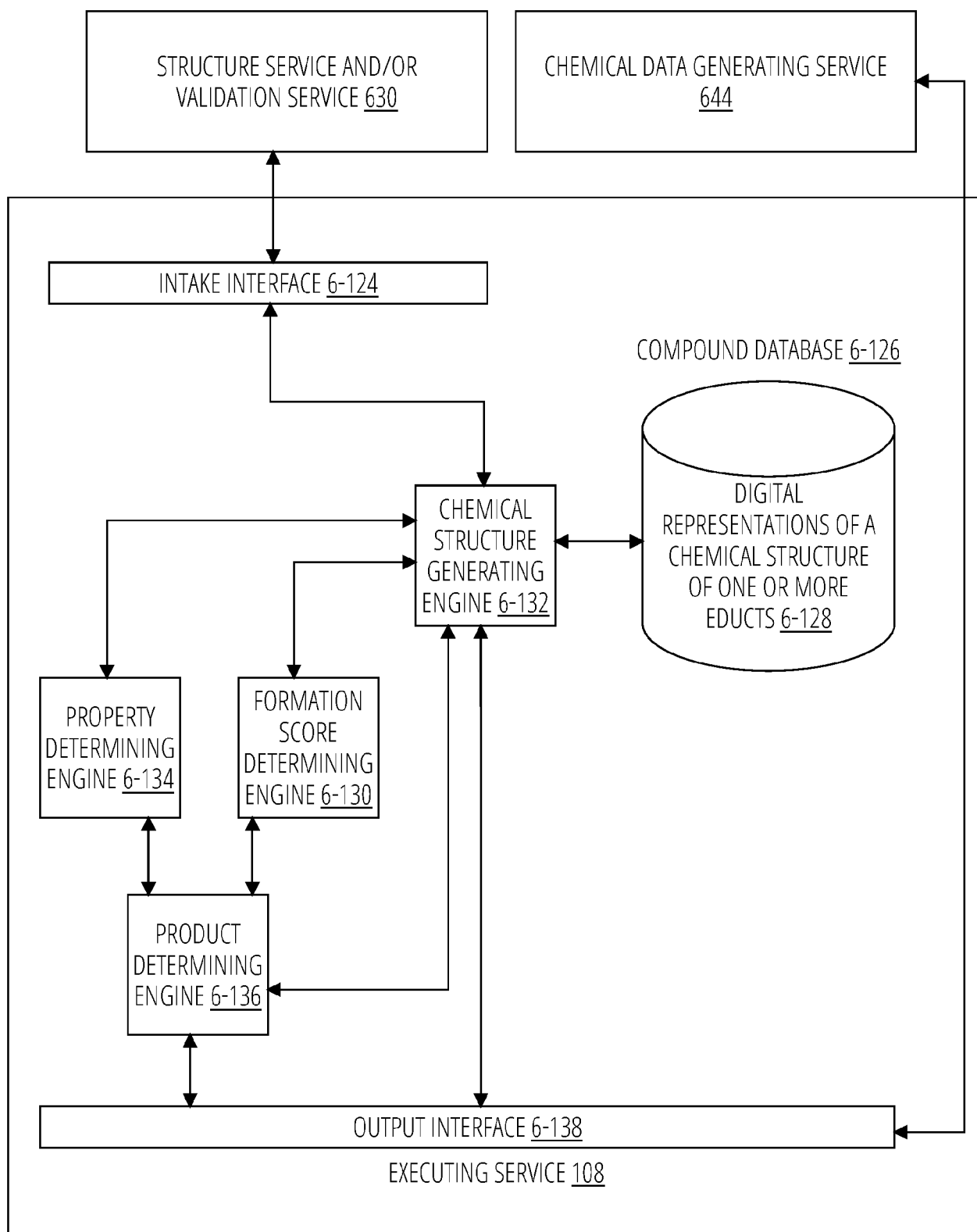
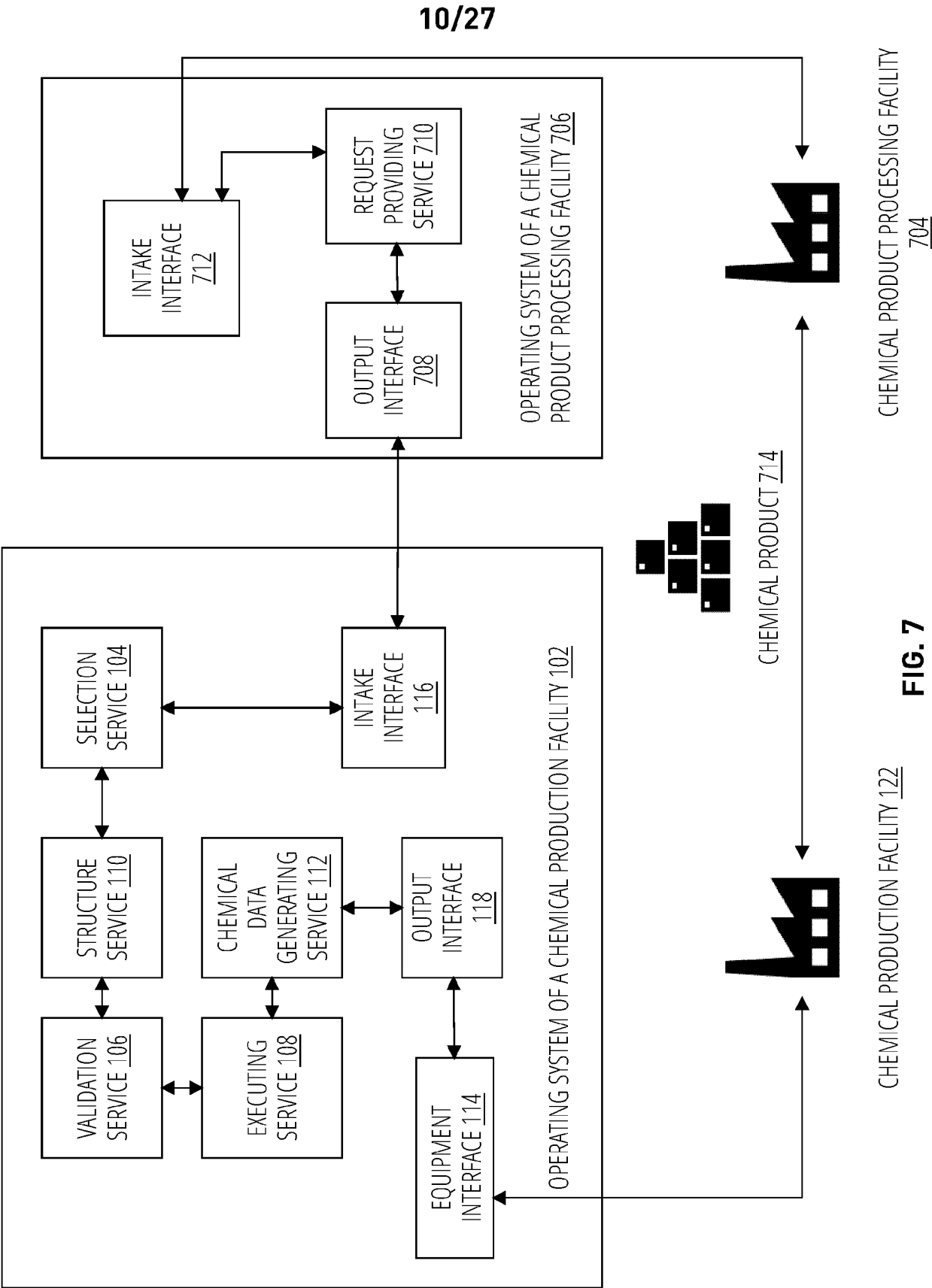


FIG. 6F



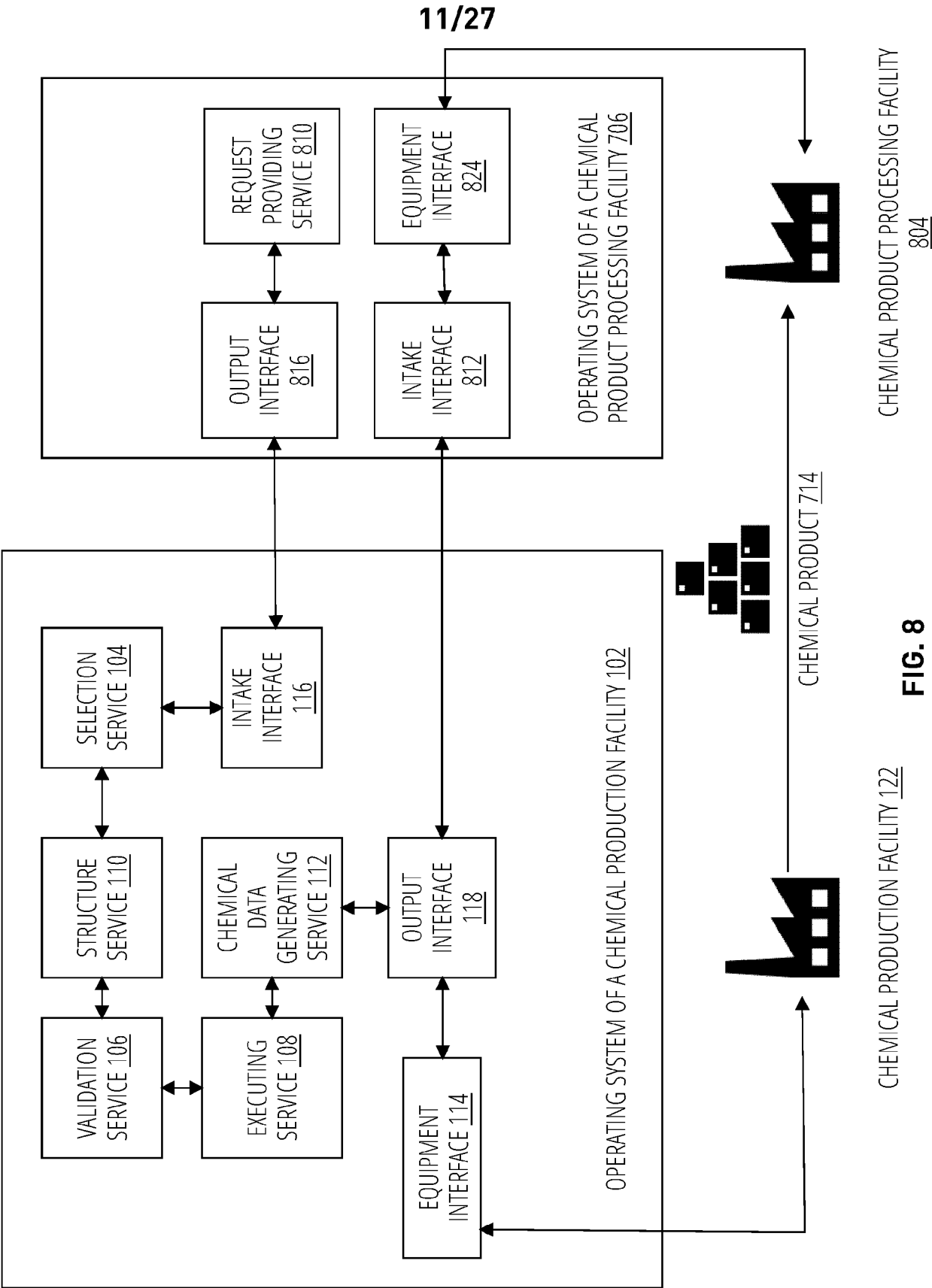


FIG. 8

12/27

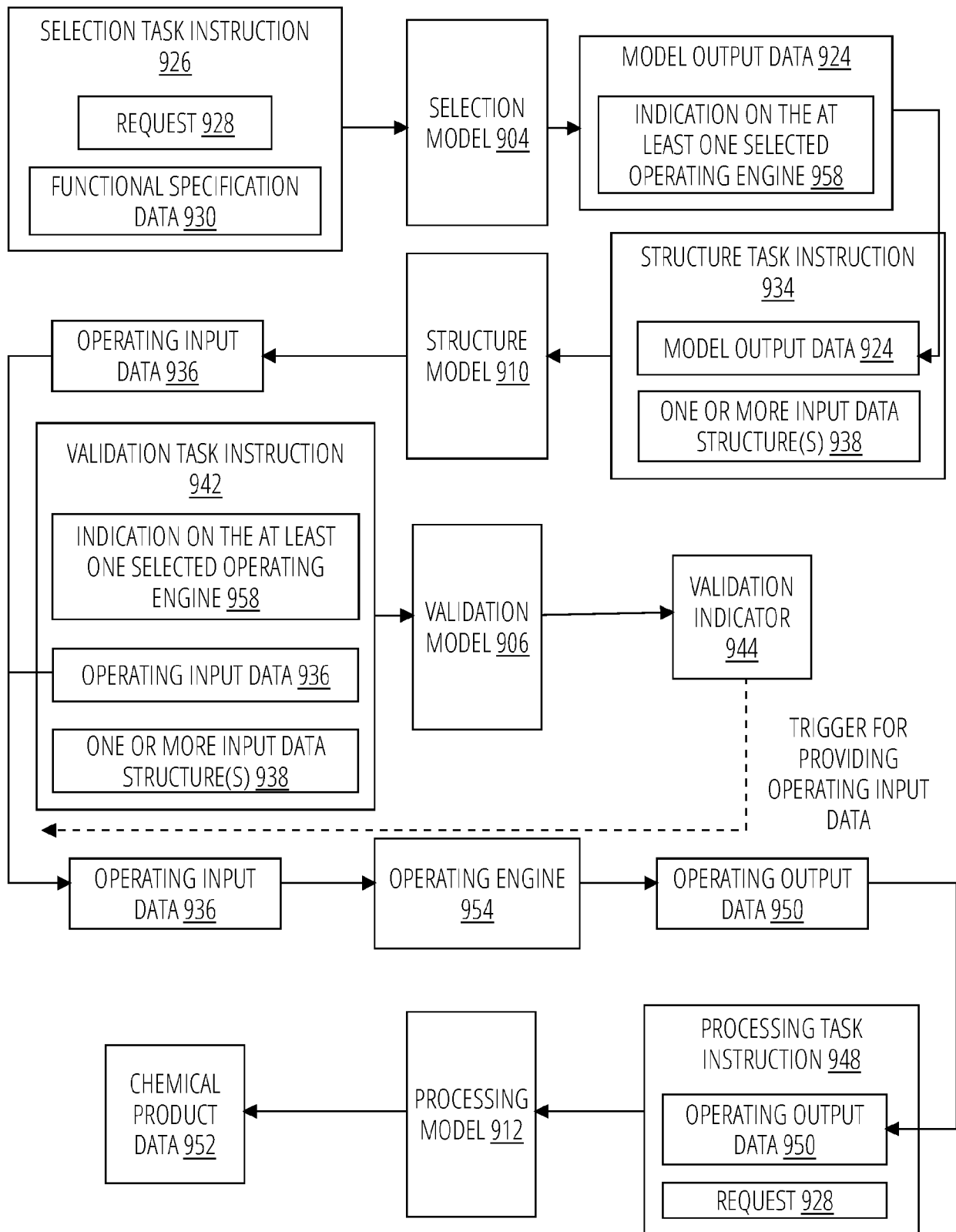


FIG. 9

13/27

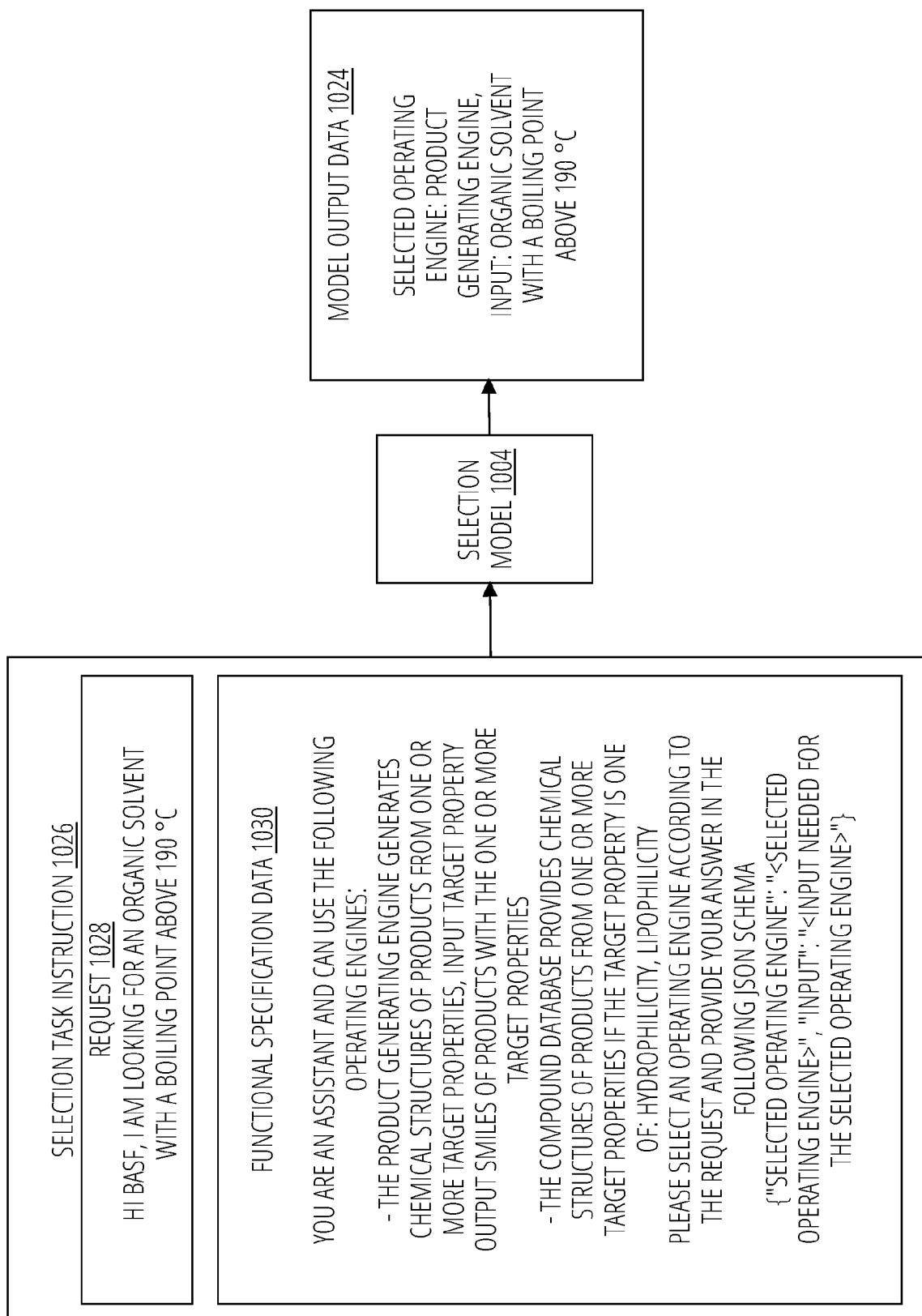


FIG. 10

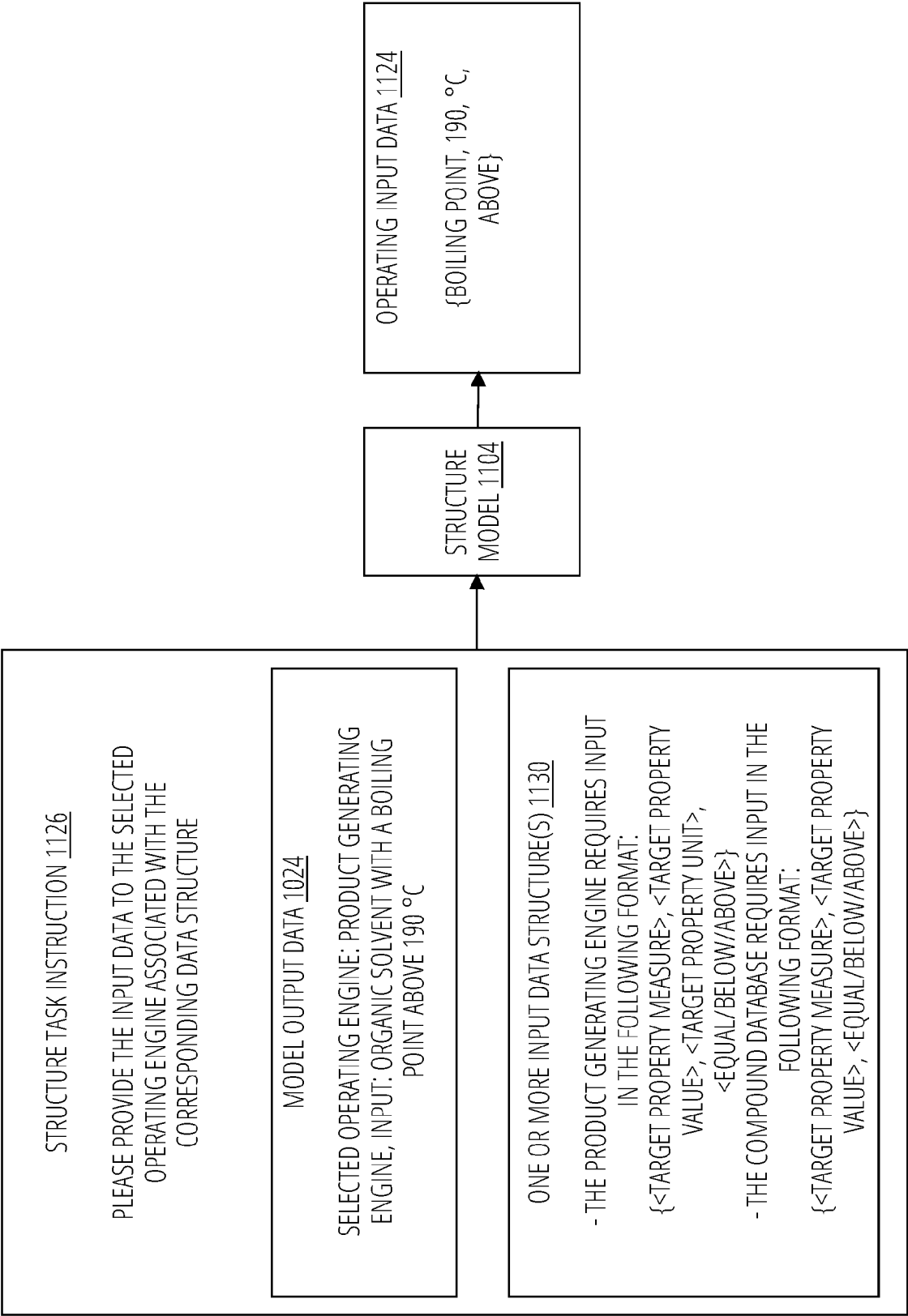


FIG. 11

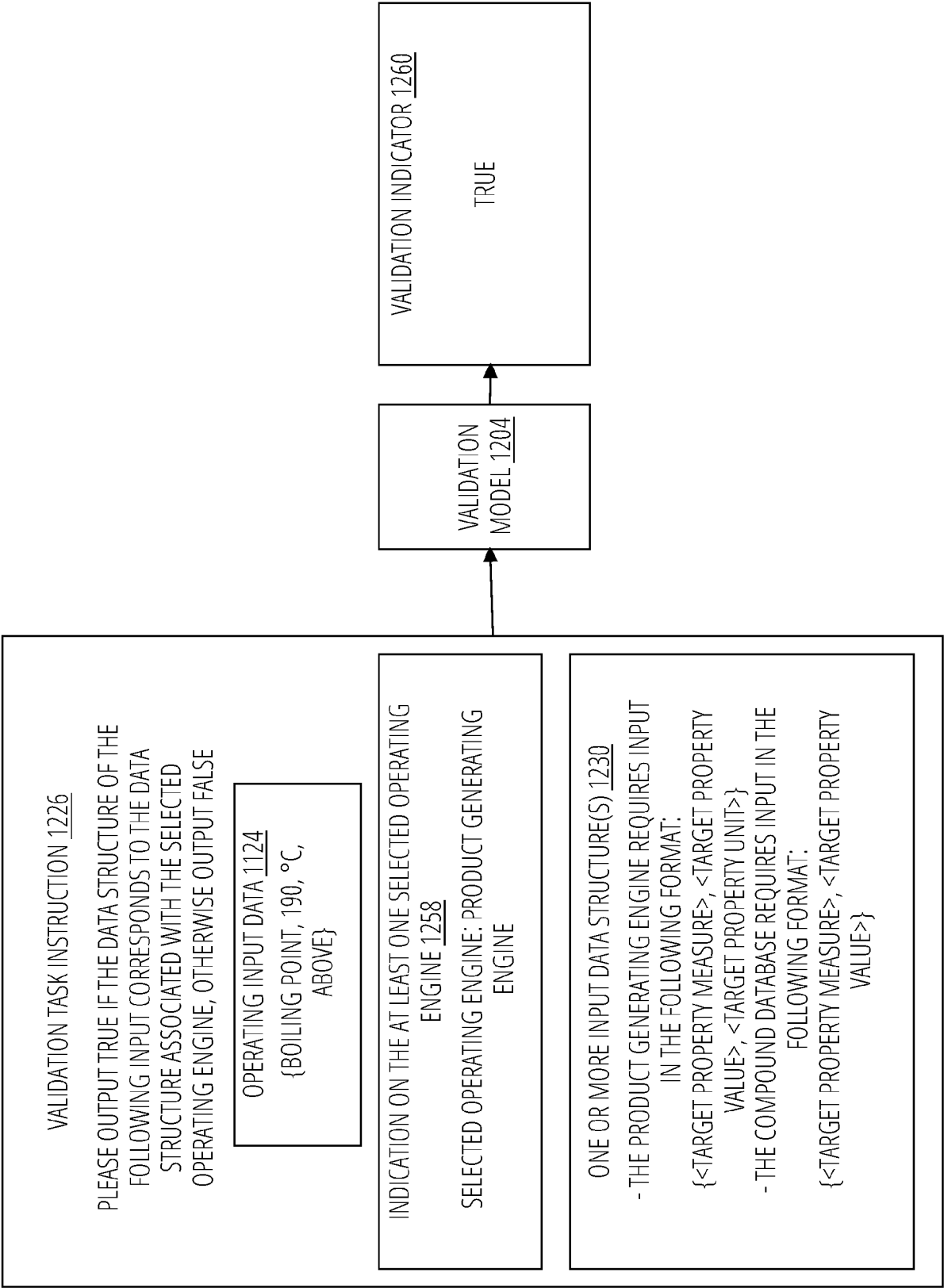


FIG. 12

16/27

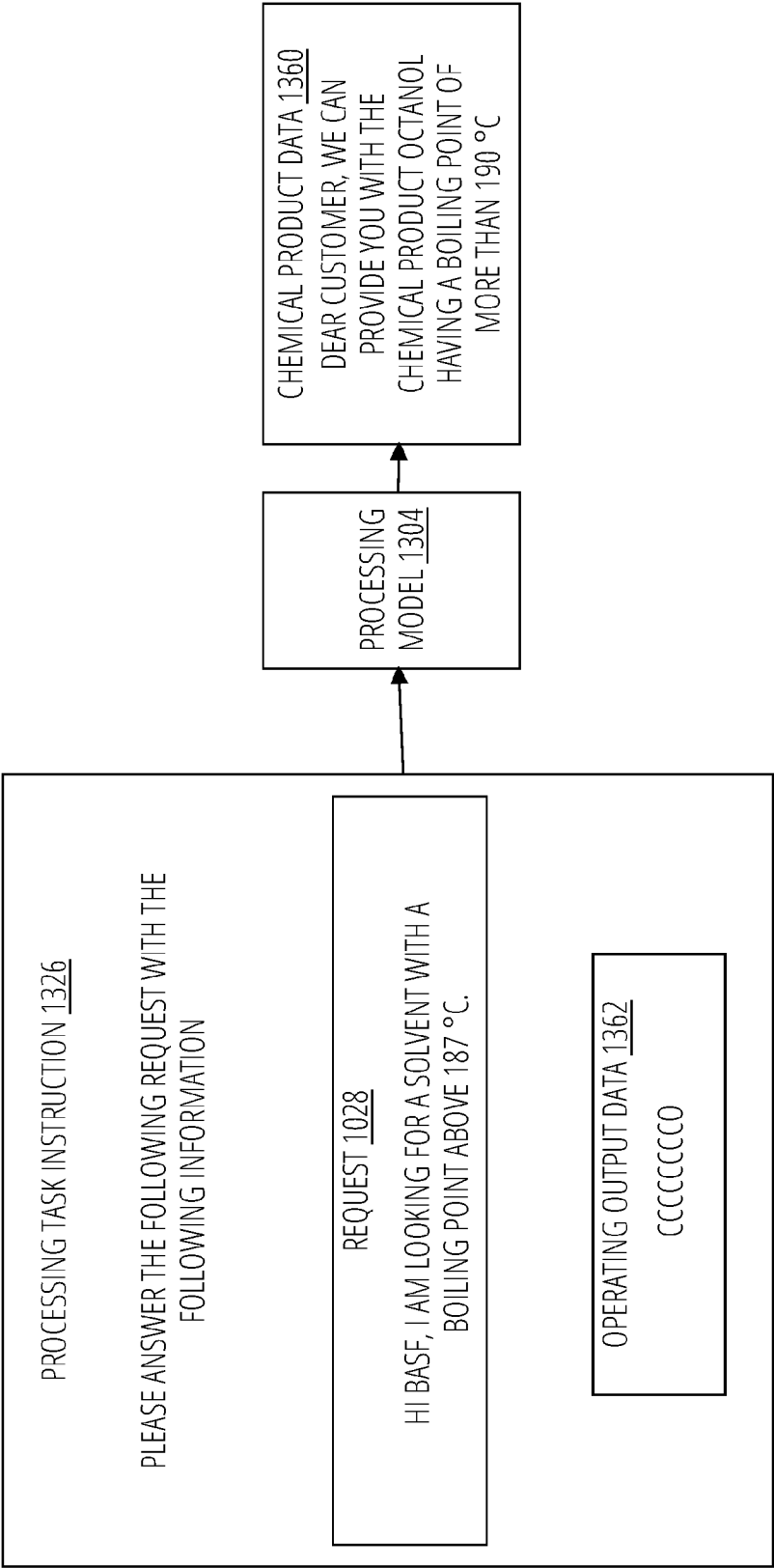


FIG. 13

17/27

REQUEST FOR RECEIVING A CHEMICAL PRODUCT WITH A TARGET PROPERTY

BASF CHEMISTRY PLATFORM	PLEASE ENTER TARGET PROPERTY OF CHEMICAL PRODUCT <u>1408</u>
	PLEASE ENTER TARGET TYPE OF CHEMICAL PRODUCT <u>1404</u>
	PLEASE ENTER TARGET FIELD OF APPLICATION ASSOCIATED WITH THE CHEMICAL PRODUCT <u>1410</u>
HI BASF, I AM LOOKING FOR A CHEMICAL THAT IS A (TYPE OF CHEMICAL PRODUCT). THE CHEMICAL PRODUCT IS TO BE APPLIED IN (FIELD OF APPLICATION). THE CHEMICAL PRODUCT SHOULD HAVE THE FOLLOWING PROPERTIES (TARGET PROPERTY). <u>1412</u>	
USER INTERFACE <u>1402</u>	

BASF CHEMISTRY PLATFORM	TARGET PROPERTY X <u>1414</u>
	TYPE OF TARGET CHEMICAL <u>1416</u>
	FIELD OF APPLICATION <u>1418</u>
DEAR CUSTOMER, WE CAN PROVIDE YOU WITH THE CHEMICAL PRODUCT X HAVING THE FOLLOWING PROPERTIES: (TARGET PROPERTY) UNTIL (DELIVERY DATE). DO YOU WANT TO ORDER? PLEASE RESPOND BY YES OR NO. <u>1420</u>	
USER INTERFACE <u>1402</u>	

FIG. 14

18/27

REQUEST FOR RECEIVING CHEMICAL PRODUCT DATA

BASF CHEMISTRY PLATFORM	PLEASE ENTER NAME OF CHEMICAL PRODUCT <u>1504</u>
	PLEASE ENTER TYPE OF CHEMICAL PRODUCT <u>1506</u>
	PLEASE ENTER FIELD OF APPLICATION ASSOCIATED WITH THE CHEMICAL PRODUCT <u>1508</u>
HI BASF, I AM LOOKING FOR PROPERTIES OF A CHEMICAL PRODUCT WITH THE NAME (NAME OF CHEMICAL PRODUCT). IT IS A (TYPE OF TARGET CHEMICAL). THE CHEMICAL IS APPLIED IN (FIELD OF APPLICATION). <u>1510</u>	
USER INTERFACE <u>1512</u>	

BASF CHEMISTRY PLATFORM	CHEMICAL PRODUCT X <u>1518</u>
	POLYMER <u>1520</u>
	SHOE SOLE <u>1522</u>
DEAR CUSTOMER, WE CAN PROVIDE YOU WITH A LIST OF PROPERTIES REGARDING THE CHEMICAL PRODUCT X. PLEASE FIND THE LIST OF PROPERTIES IN THE DOCUMENT HERE: /PROPERTY_DOCUMENT_CHEMICAL_PRODUCT_X/ <u>1524</u>	
USER INTERFACE <u>1512</u>	

FIG. 15

19/27

REQUEST FOR RECEIVING A REPRESENTATION OF A CHEMICAL PRODUCT

BASF CHEMISTRY PLATFORM	PLEASE ENTER NAME OF CHEMICAL PRODUCT <u>1604</u>
PLEASE UPLAD IMAGE OF CHEMICAL PRODUCT <u>1628</u>	PLEASE ENTER TYPE OF CHEMICAL PRODUCT <u>1606</u>
	PLEASE ENTER FIELD OF APPLICATION ASSOCIATED WITH THE CHEMICAL PRODUCT <u>1608</u>
	PLEASE ENTER ONE OR MORE INGREDIENTS <u>1624</u>
HI BASF, DOES THE PRODUCT (CHEMICAL PRODUCT) SPECIFIED BY (TYPE OF CHEMICAL PRODUCT, FIELD OF APPLICATION, IMAGE) COMPRISE ONE ORF THE FOLLOWING INGREDIENTS: (ONE OR MORE INGREDIENTS)? <u>1610</u>	
USER INTERFACE <u>1612</u>	

BASF CHEMISTRY PLATFORM	GLYSANTIN <u>1616</u>
/IMAGE OF CHEMICAL PRODUCT/ <u>1630</u>	LIQUID <u>1618</u>
	COOLANT ADDITIVE <u>1620</u>
	MONOETHYLENGLYCOL AND INHIBITOR A <u>1626</u>
DEAR CUSTOMER, THE CHEMICAL PRODUCT COMPRISES MONOTHYLENGLYCOL AND INHIBITOR B AND C. DO YOU WANT MORE INFORMATION ON THE INGREDIENTS OF GLYSANTIN? <u>1622</u>	
USER INTERFACE <u>1612</u>	

FIG. 16

20/27

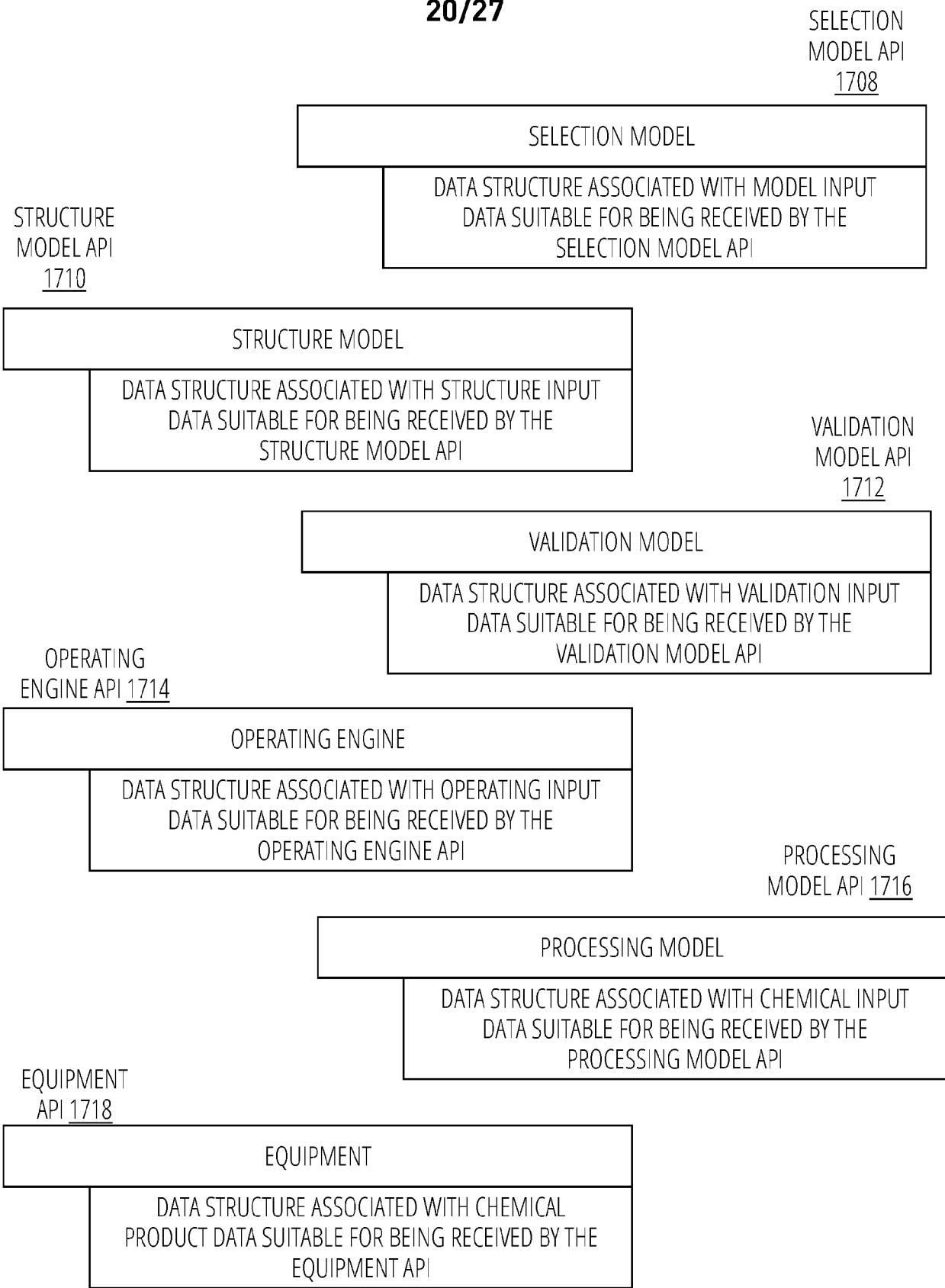


FIG. 17

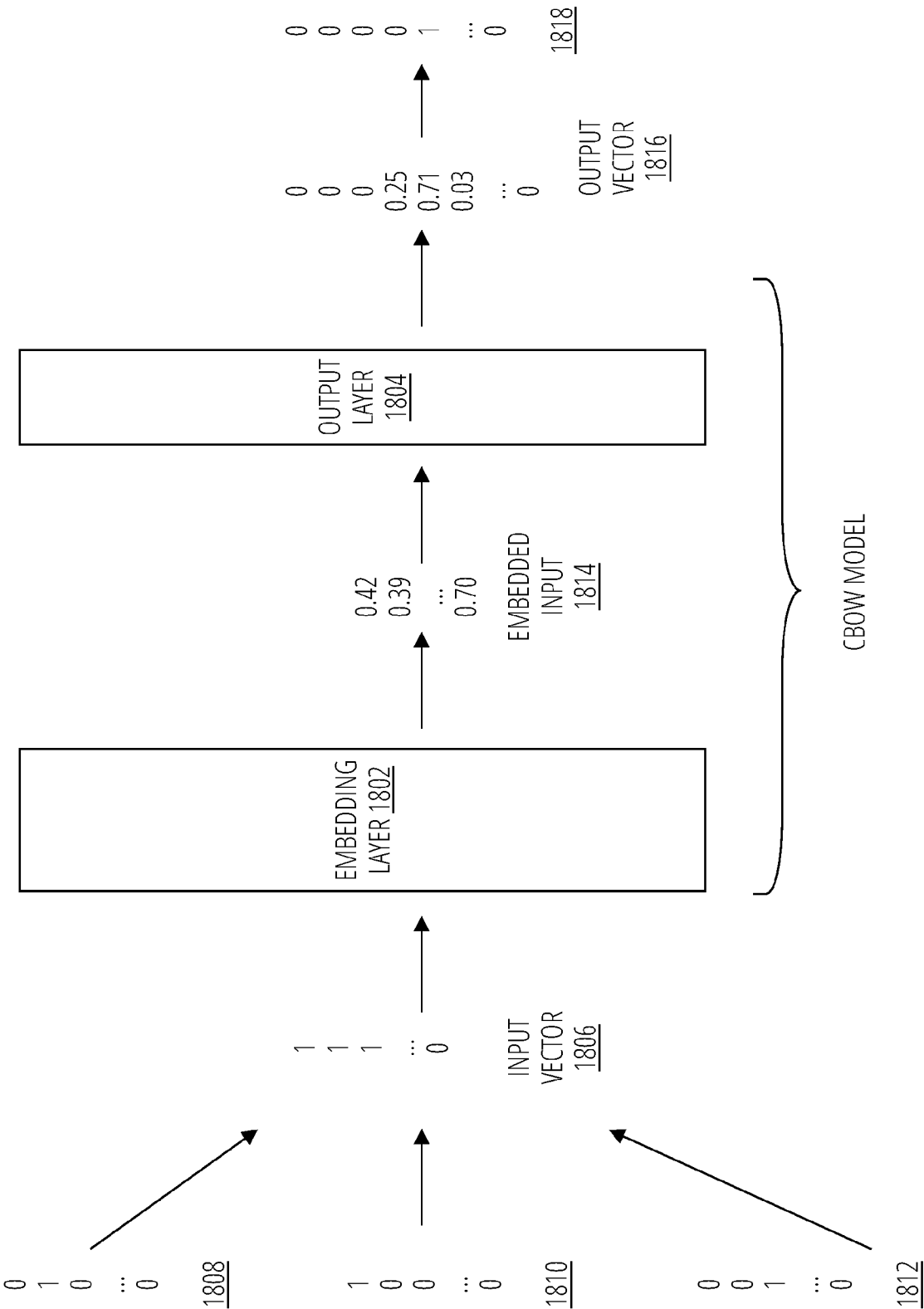


FIG. 18

22/27

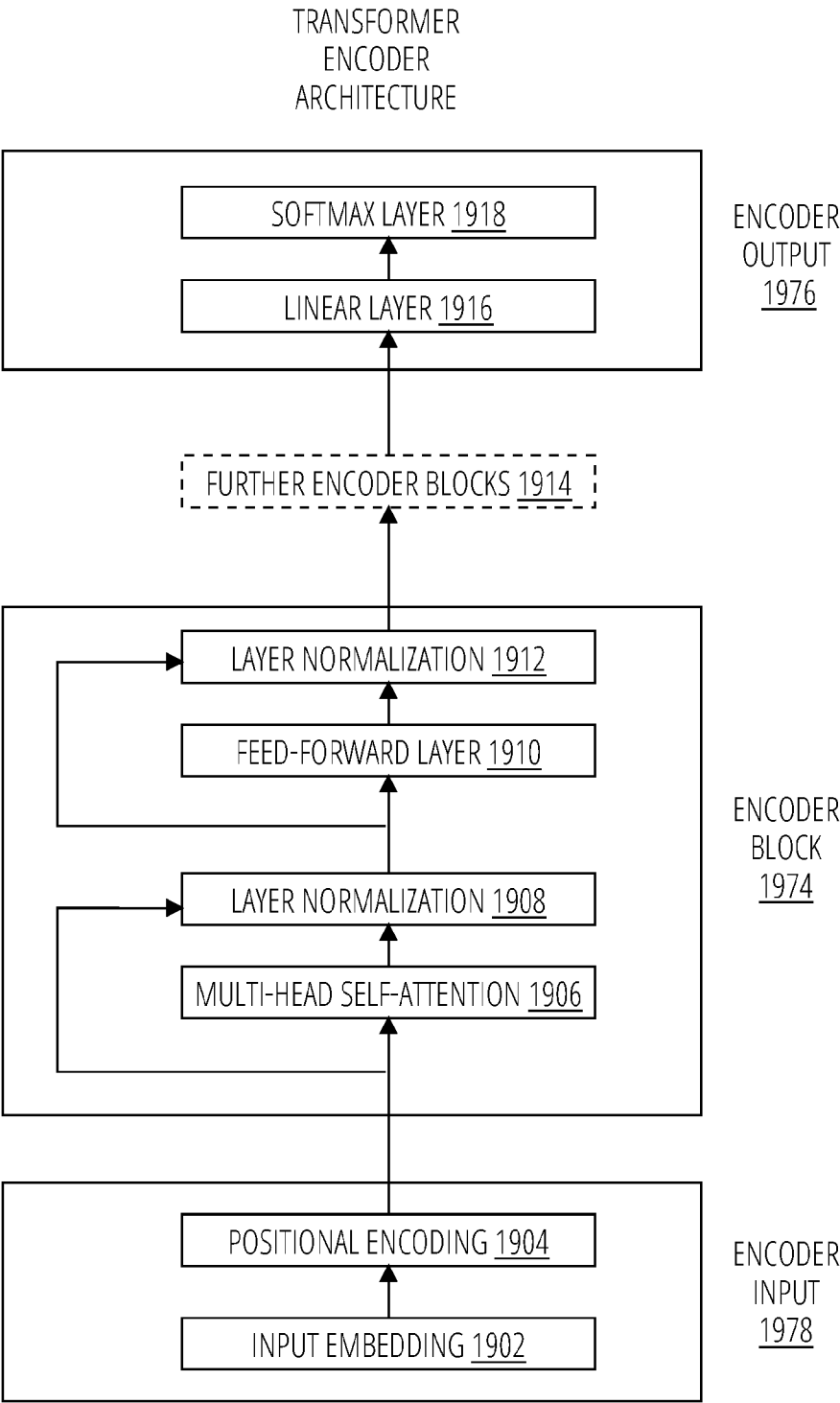


FIG. 19A

23/27

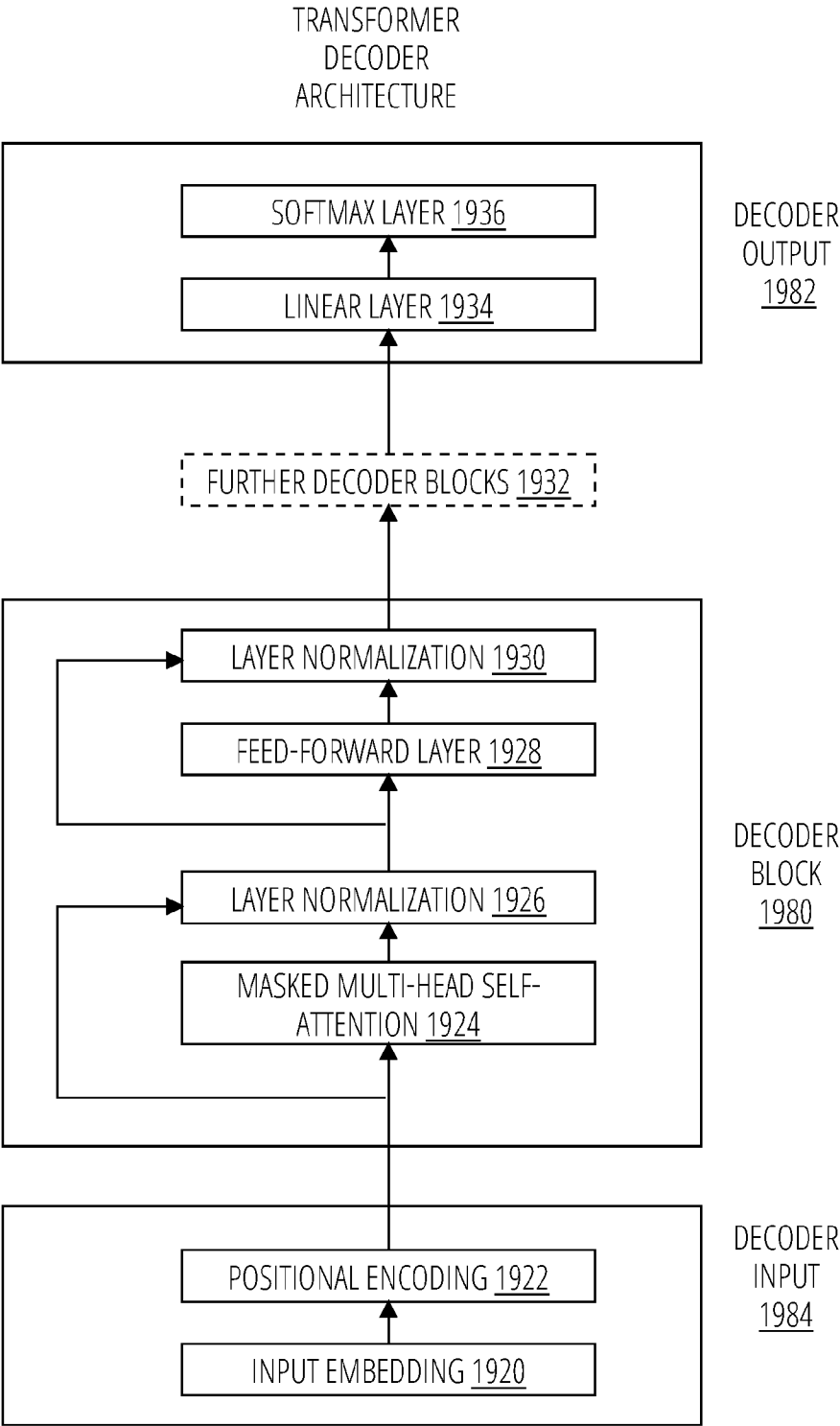


FIG. 19B

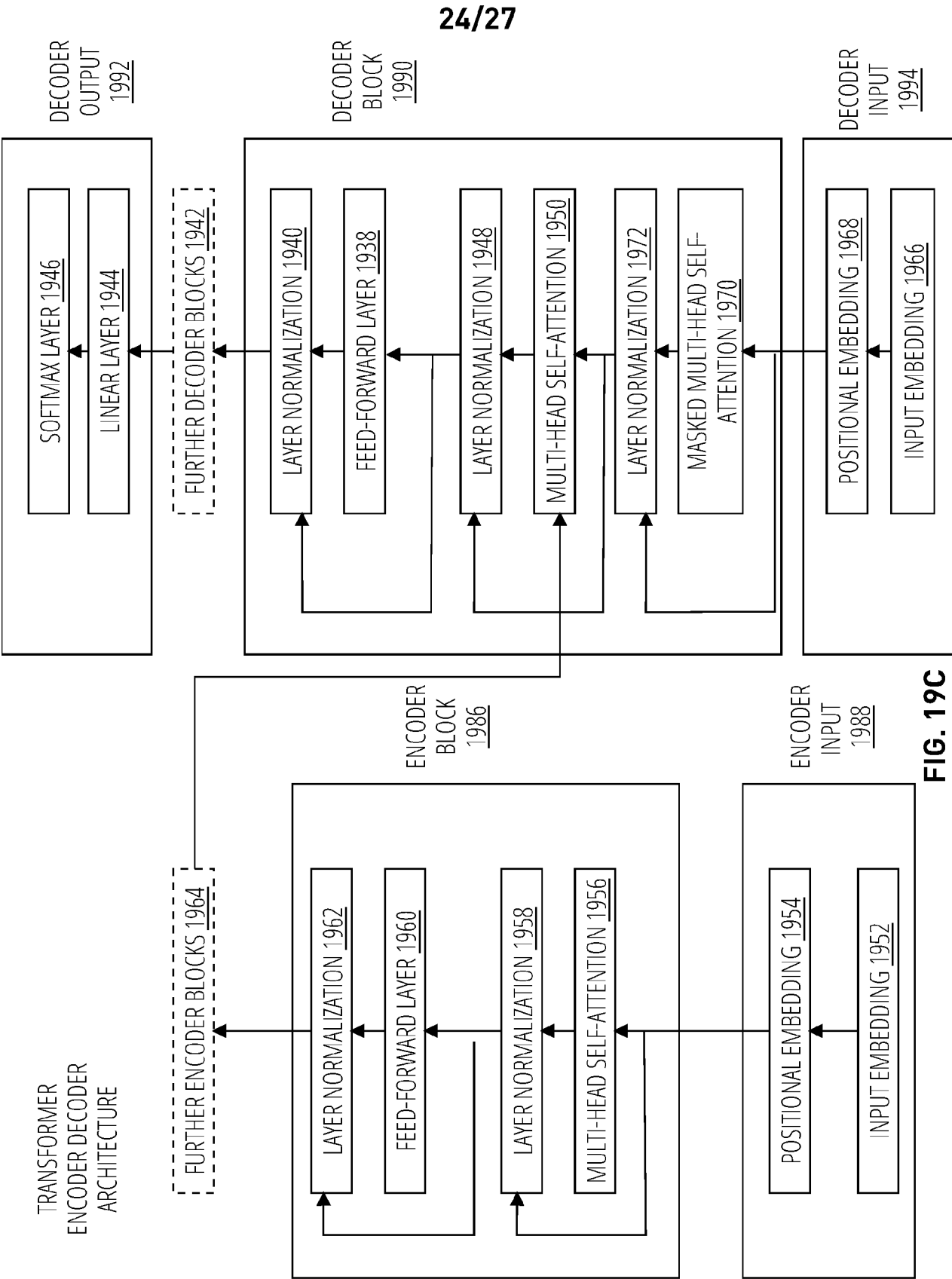


FIG. 19C

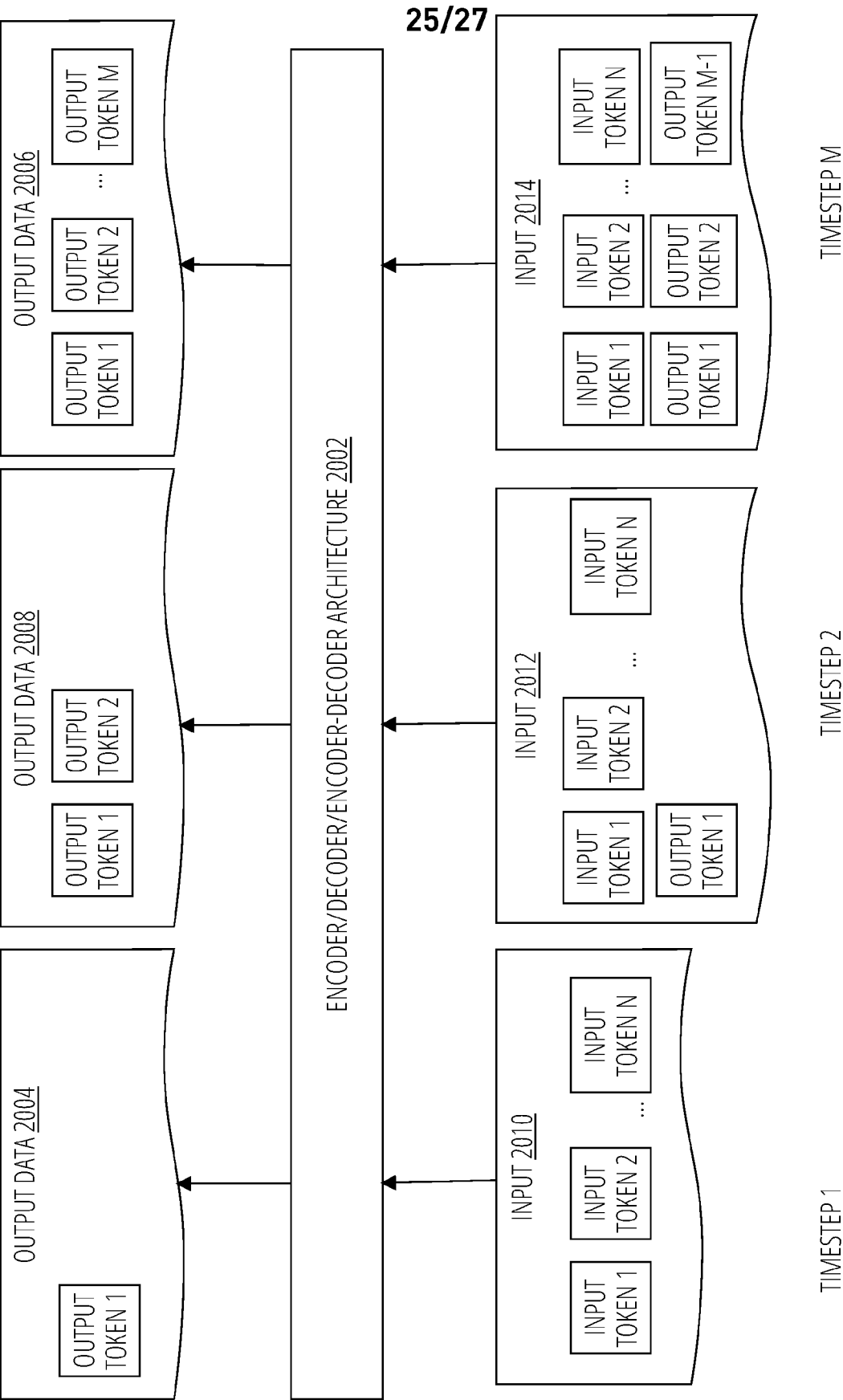


FIG. 20

TABLE 2102	
COL 1	COL 2
1	2
3	4

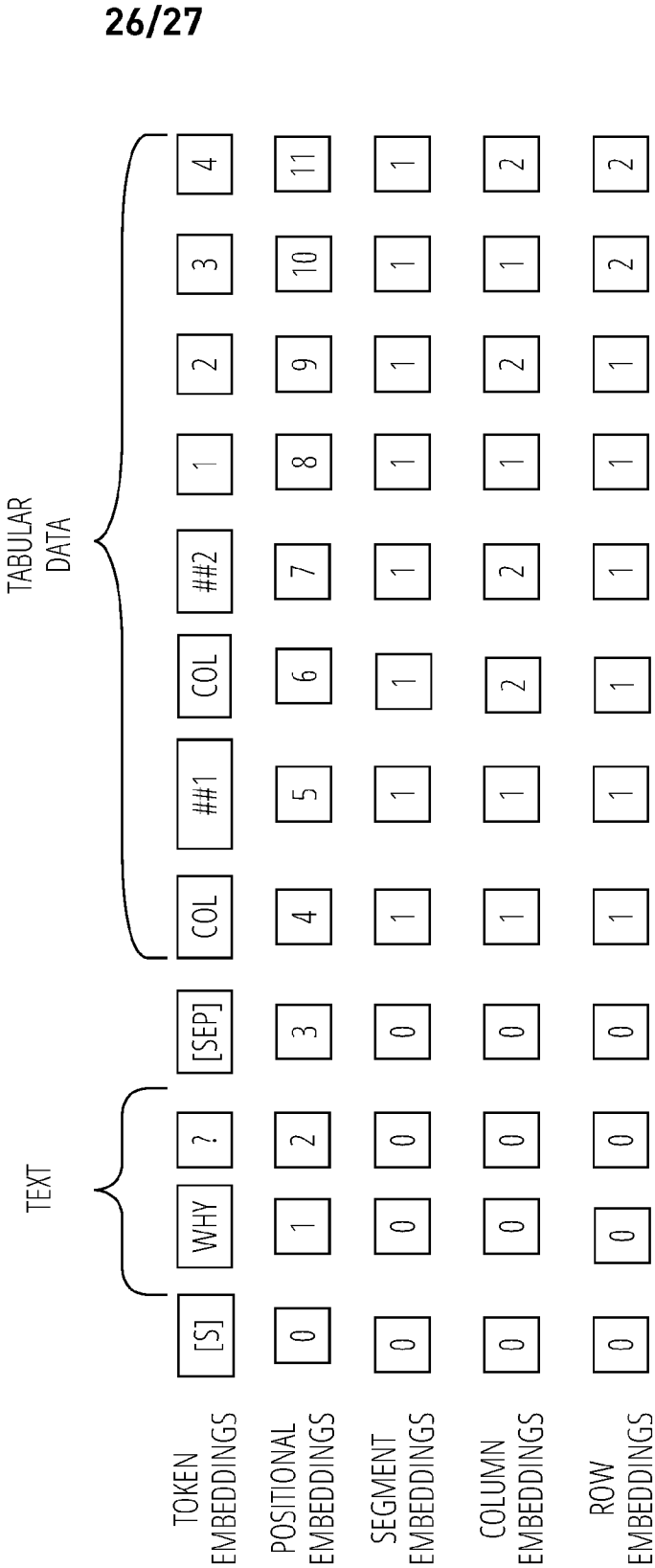


FIG. 21

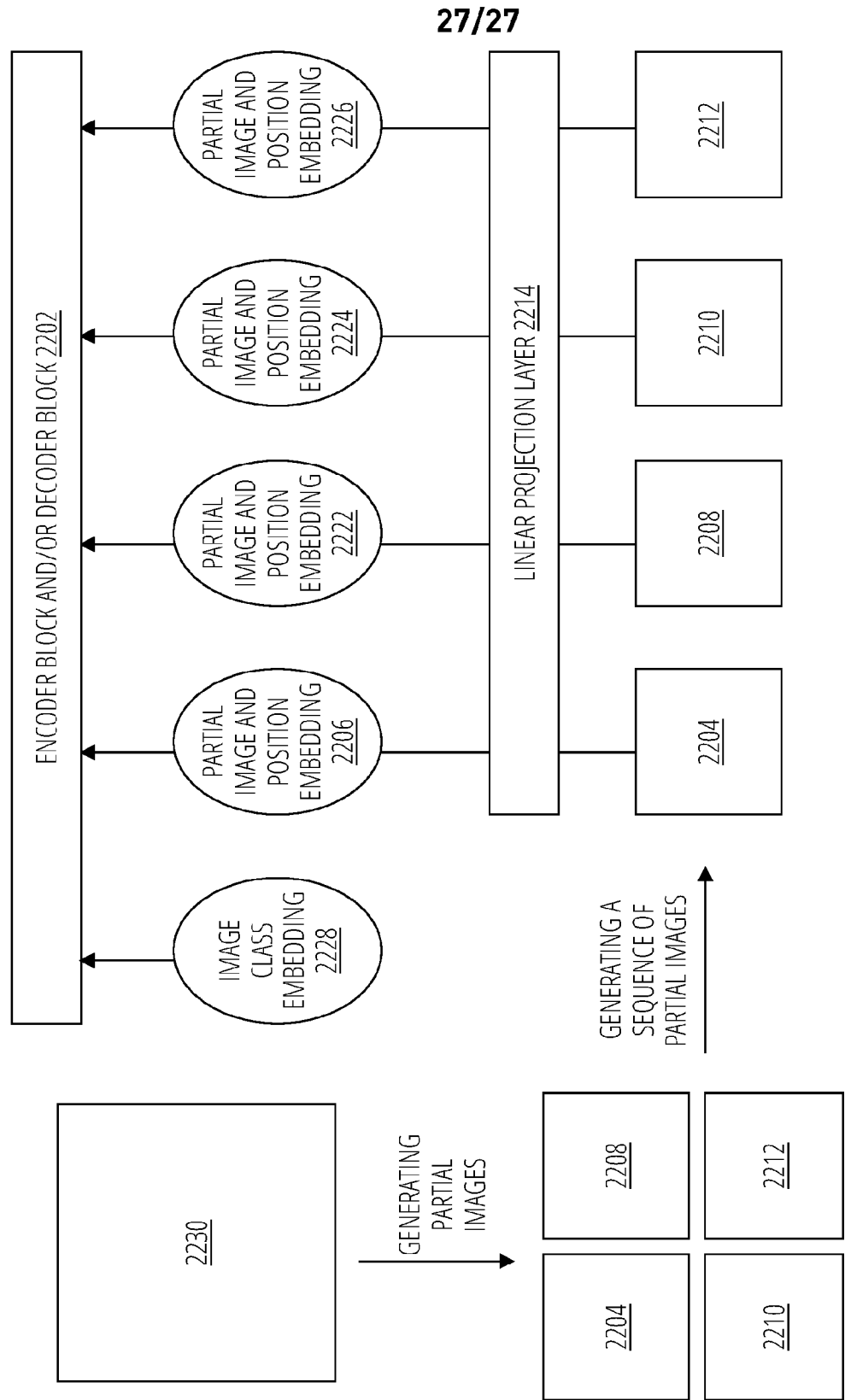


FIG. 22

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2024/085680

A. CLASSIFICATION OF SUBJECT MATTER INV. G06Q10/0631 G06Q50/04 ADD. G06N3/02		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G06Q G06N G06F Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2021/280276 A1 (GUPTA VIVEK [IN] ET AL) 9 September 2021 (2021-09-09) figures 1-4 paragraph [0018] - paragraph [0026] paragraph [0053] - paragraph [0056] -----	1 - 15
X	WO 2023/094667 A1 (BASF SE [DE]) 1 June 2023 (2023-06-01) figures 4, 5 page 8, line 28 - line 30 page 10, line 35 page 11, line 17 page 12, line 1 - line 3 page 13, line 10 - page 15, line 26 page 16, line 26 - page 17, line 32 page 18, line 36 - page 19, line 26 page 20, line 31 - line 34 ----- - / - -	1 - 15
☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 10 March 2025		Date of mailing of the international search report 20/03/2025
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Komnios, Ioannis

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2024/085680

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2023/341850 A1 (CELLA CHARLES H [US] ET AL) 26 October 2023 (2023-10-26) figure 71 paragraph [0040] paragraph [0554] paragraph [1074] - paragraph [1077] paragraph [1085] paragraph [1100] - paragraph [1101] paragraph [1270] -----	1 - 15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2024/085680

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2021280276 A1	09-09-2021	EP 3896699 A1	20-10-2021
		US 2021280276 A1	09-09-2021

WO 2023094667 A1	01-06-2023	EP 4436374 A1	02-10-2024
		WO 2023094667 A1	01-06-2023

US 2023341850 A1	26-10-2023	AU 2021401816 A1	22-06-2023
		EP 4264518 A1	25-10-2023
		JP 2024500818 A	10-01-2024
		KR 20230135069 A	22-09-2023
		US 2022193915 A1	23-06-2022
		US 2022196889 A1	23-06-2022
		US 2022197246 A1	23-06-2022
		US 2022197247 A1	23-06-2022
		US 2022197306 A1	23-06-2022
		US 2022305735 A1	29-09-2022
		US 2022318707 A1	06-10-2022
		US 2023089330 A1	23-03-2023
		US 2023090334 A1	23-03-2023
		US 2023097438 A1	30-03-2023
		US 2023103078 A1	30-03-2023
		US 2023135553 A1	04-05-2023
		US 2023339108 A1	26-10-2023
		US 2023339114 A1	26-10-2023
		US 2023341850 A1	26-10-2023
		US 2023342684 A1	26-10-2023
