Large Process Models

Peter Fettke

Deutsches Forschungszentrum für Künstliche Intelligenz
German Research Center for Artificial Intelligence

UNIVERSITÄT DES SAARLANDES

18th Symposium and Summer School
on Service-Oriented Computing (SummerSoC)
Crete, Greece
June 24th - 29th, 2024

Large Process Models

Peter Fettke

in cooperation with Wolfgang Reisig

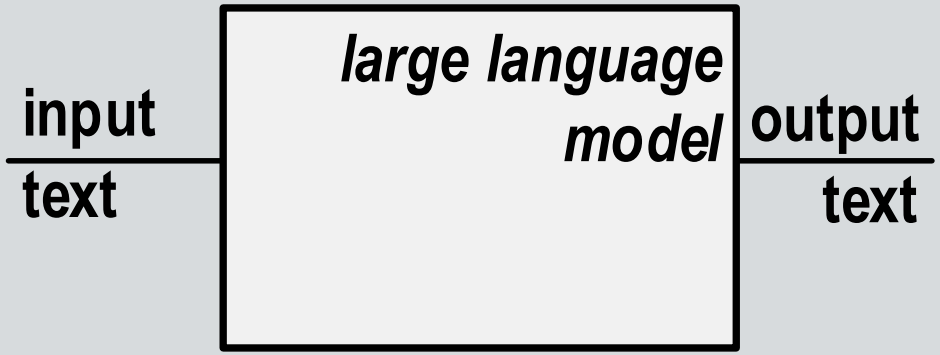
Deutsches Forschungszentrum für Künstliche Intelligenz
German Research Center for Artificial Intelligence

UNIVERSITÄT DES SAARLANDES

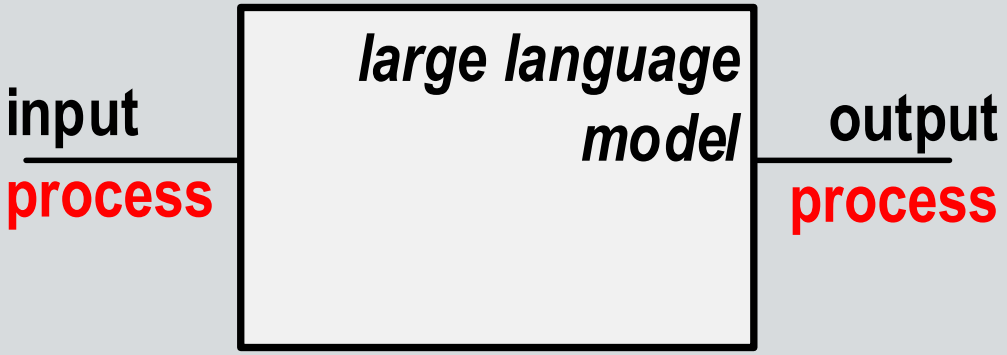
HUMBOLDT-UNIVERSITÄT ZU BERLIN

18th Symposium and Summer School
on Service-Oriented Computing (SummerSoC)
Crete, Greece
June 24th - 29th, 2024

This talk in a nutshell



This talk in a nutshell



Agenda

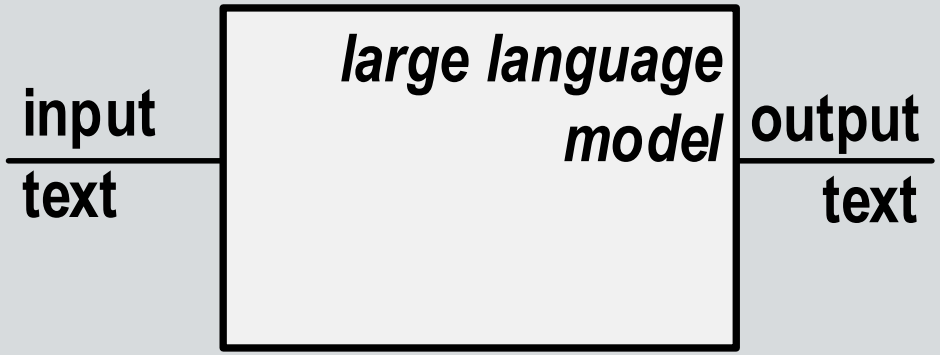
1. Large Language Models
2. Process Models
3. Large Process Models

dfki ai Deutsches Forschungszentrum für Künstliche Intelligenz
German Research Center for Artificial Intelligence

UNIVERSITÄT DES SAARLANDES

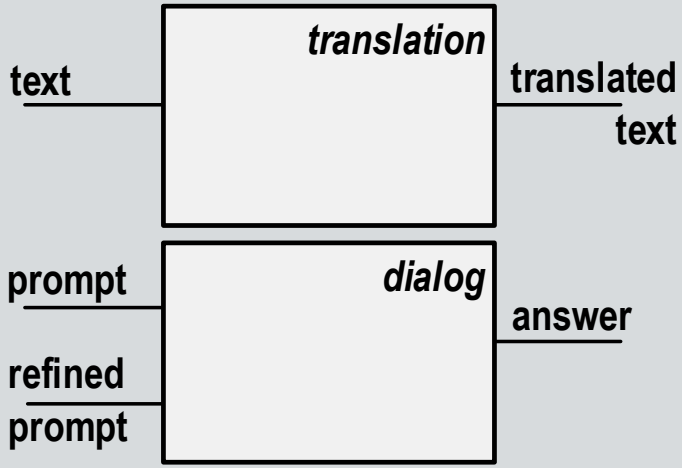
HUMBOLDT-UNIVERSITÄT ZU BERLIN

What is a large language model?
Most abstract view

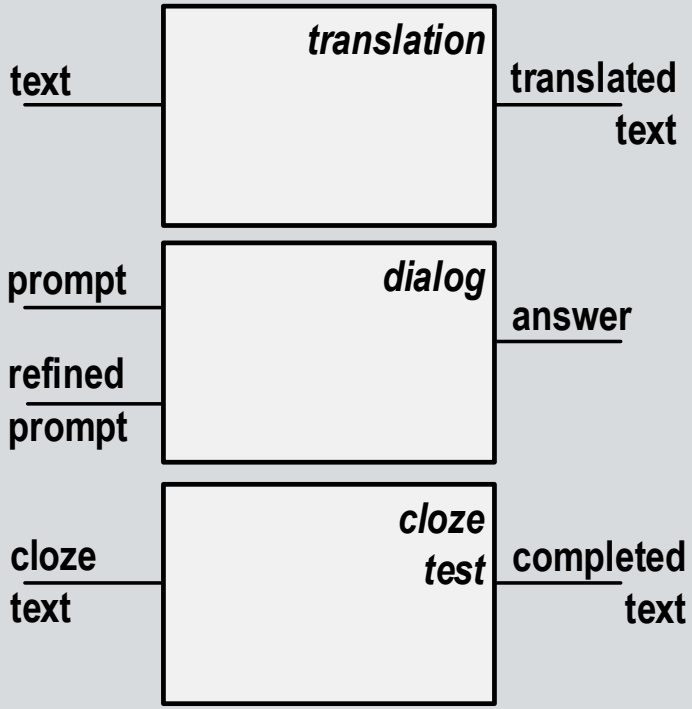


What is a large language model? More details

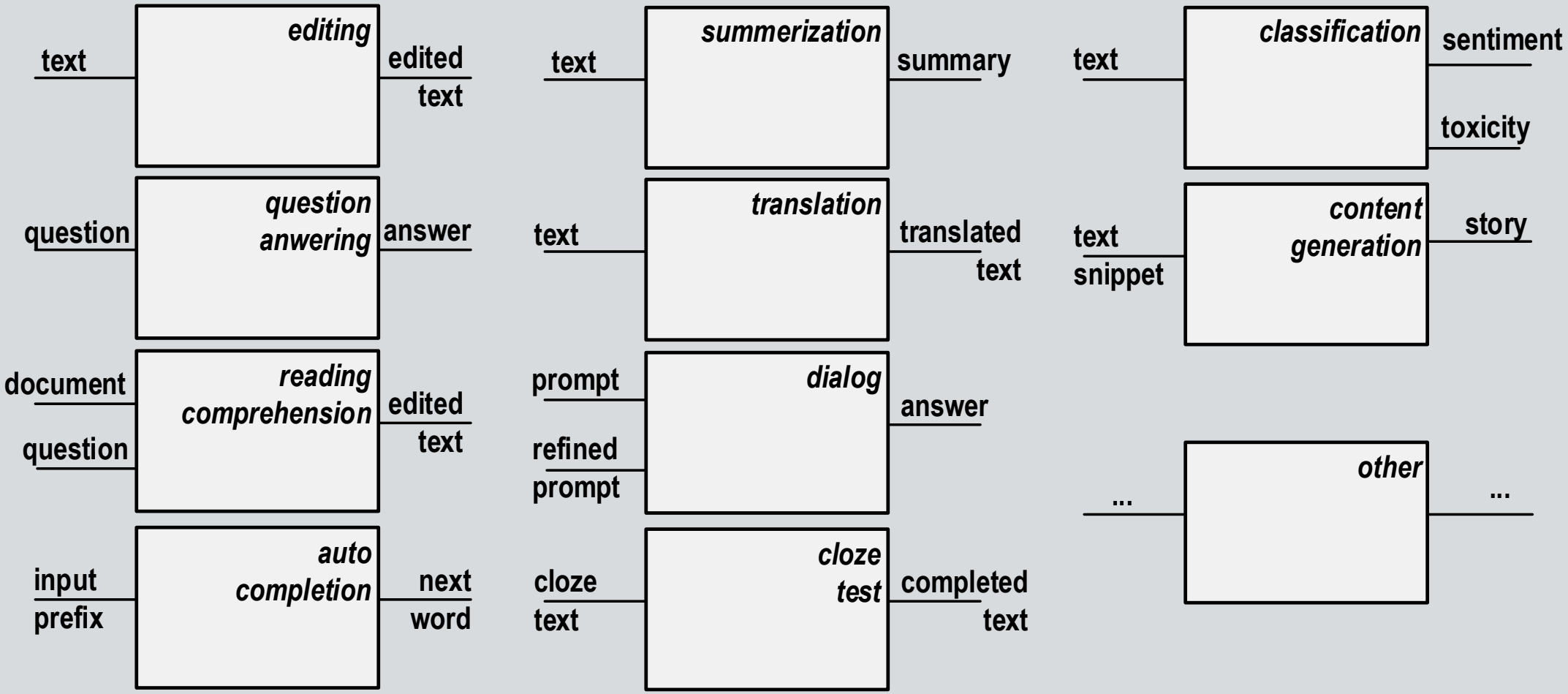
What is a large language model? More details



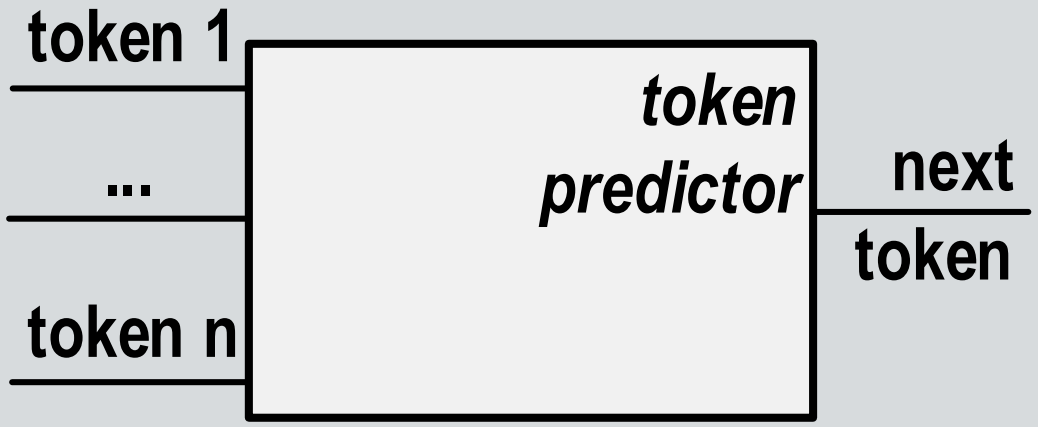
What is a large language model? More details



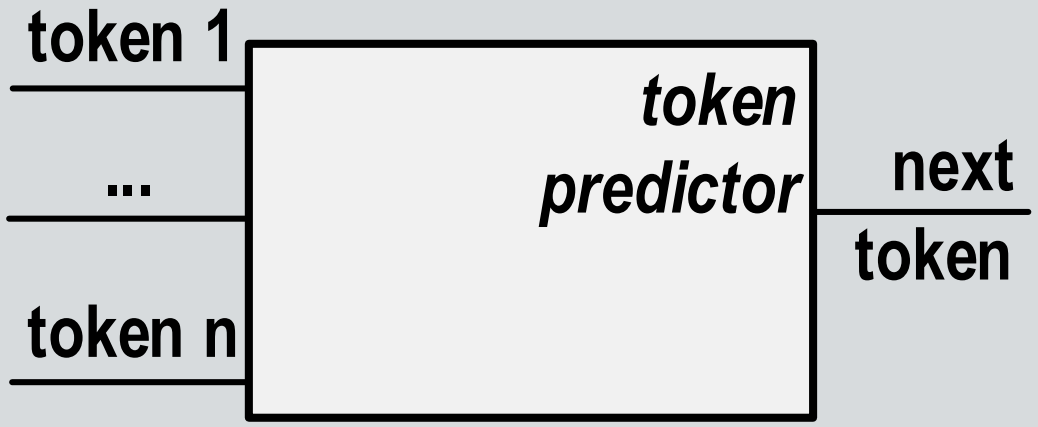
What is a large language model? More details



What is a large language model? A technical view



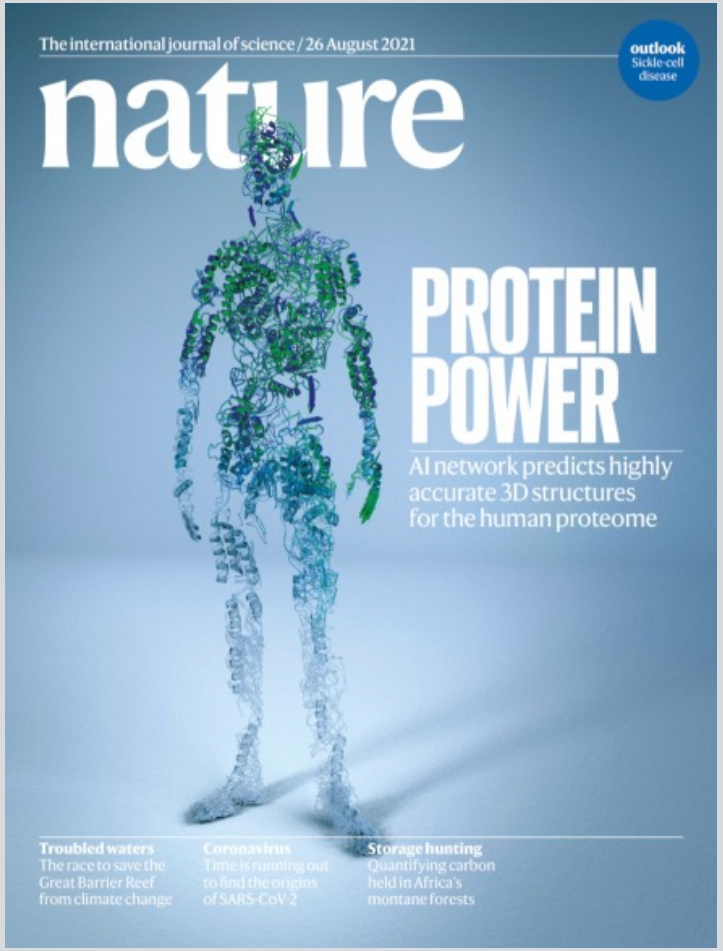
What is a large language model? A technical view



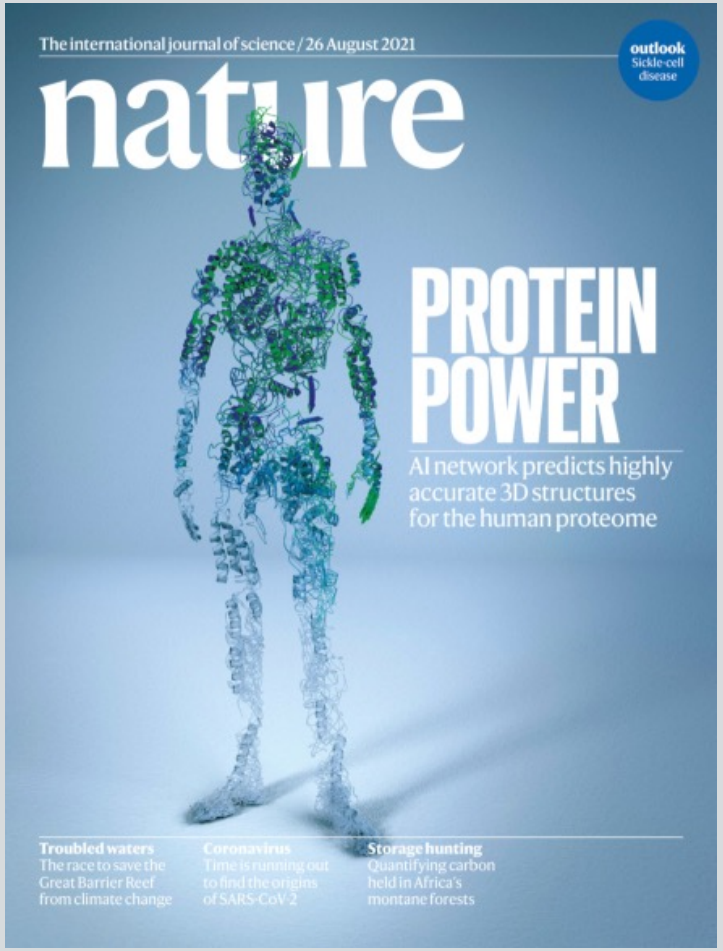
What is a token?

- character
- word
- n-gram
- part of a word
- byte pair encoding (BPE)

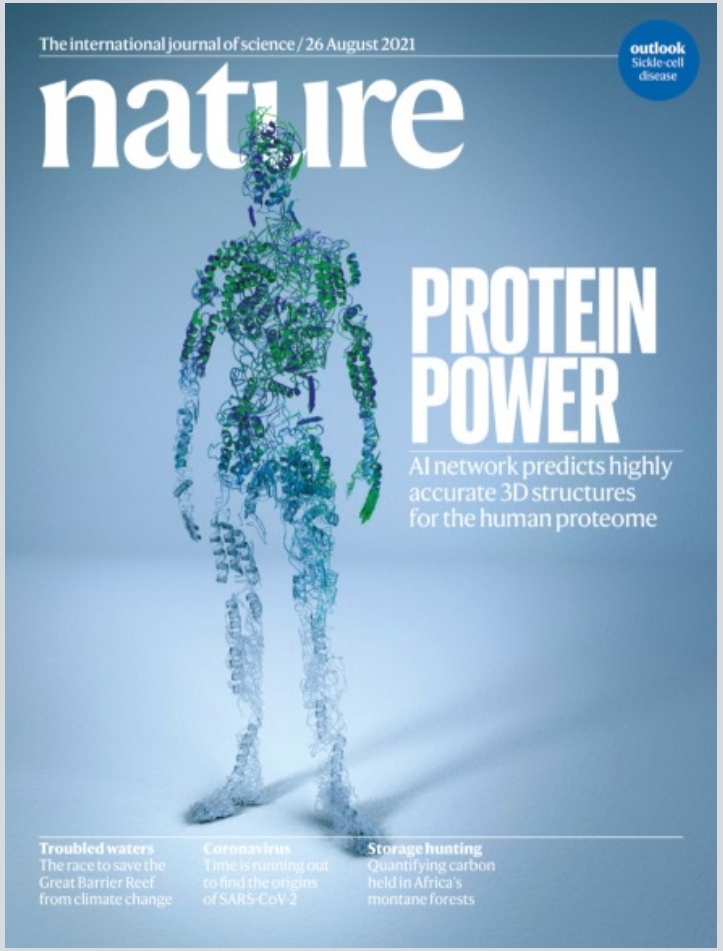
Large language models in non-text domains



Large language models in non-text domains



Large language models in non-text domains



Article

Accurate structure prediction of biomolecular interactions with AlphaFold 3

https://doi.org/10.1038/s41586-024-07487-w
Received: 19 December 2023
Accepted: 29 April 2024
Published online: 8 May 2024
Open access
Check for updates

Josh Abramson[1,7], Jonas Adler[1,7], Jack Dunger[1,7], Richard Evans[1,7], Tim Green[1,7], Alexander Pritzel[1,7], Olaf Ronneberger[1,7], Lindsay Willmore[1,7], Andrew J. Ballard[1], Joshua Bambrick[2], Sebastian W. Bodenstein[1], David A. Evans[1], Chia-Chun Hung[2], Michael O'Neill[1], David Reiman[1], Kathryn Tunyasuvunakool[1], Zachary Wu[1], Akvilė Žemgulytė[1], Eirini Arvaniti[3], Charles Beattie[3], Ottavia Bertolli[3], Alex Bridgland[3], Alexey Cherepanov[4], Miles Congreve[4], Alexander I. Cowen-Rivers[3], Andrew Cowie[3], Michael Figurnov[3], Fabian B. Fuchs[3], Hannah Gladman[3], Rishub Jain[3], Yousuf A. Khan[3,5], Caroline M. R. Low[4], Kuba Perlin[3], Anna Potapenko[3], Pascal Savy[4], Sukhdeep Singh[3], Adrian Stecula[4], Ashok Thillaisundaram[3], Catherine Tong[4], Sergei Yakneen[4], Ellen D. Zhong[3,6], Michal Zielinski[3], Augustin Žídek[3], Victor Bapst[1,8], Pushmeet Kohli[1,8], Max Jaderberg[2,8] ✉, Demis Hassabis[1,2,8] ✉ & John M. Jumper[1,8] ✉

The introduction of AlphaFold 2[1] has spurred a revolution in modelling the structure of proteins and their interactions, enabling a huge range of applications in protein modelling and design[2–6]. Here we describe our AlphaFold 3 model with a substantially updated diffusion-based architecture that is capable of predicting the joint structure of complexes including proteins, nucleic acids, small molecules, ions and modified residues. The new AlphaFold model demonstrates substantially improved accuracy over many previous specialized tools: far greater accuracy for protein–ligand interactions compared with state-of-the-art docking tools, much higher accuracy for protein–nucleic acid interactions compared with nucleic-acid-specific predictors and substantially higher antibody–antigen prediction accuracy compared with AlphaFold-Multimer v.2.3[7,8]. Together, these results show that high-accuracy modelling across biomolecular space is possible within a single unified deep-learning framework.

Accurate models of biological complexes are critical to our understanding of cellular functions and for the rational design of therapeutics[1–4]. Enormous progress has been achieved in protein structure prediction with the development of AlphaFold[1], and the field has grown tremendously with a number of later methods that build on the ideas and techniques of AlphaFold 2 (AF2)[10–12]. Almost immediately after AlphaFold became available, it was shown that simple input modifications would enable surprisingly accurate protein interaction predictions[13–15] and that training AF2 specifically for protein interaction prediction yielded a highly accurate system[7].

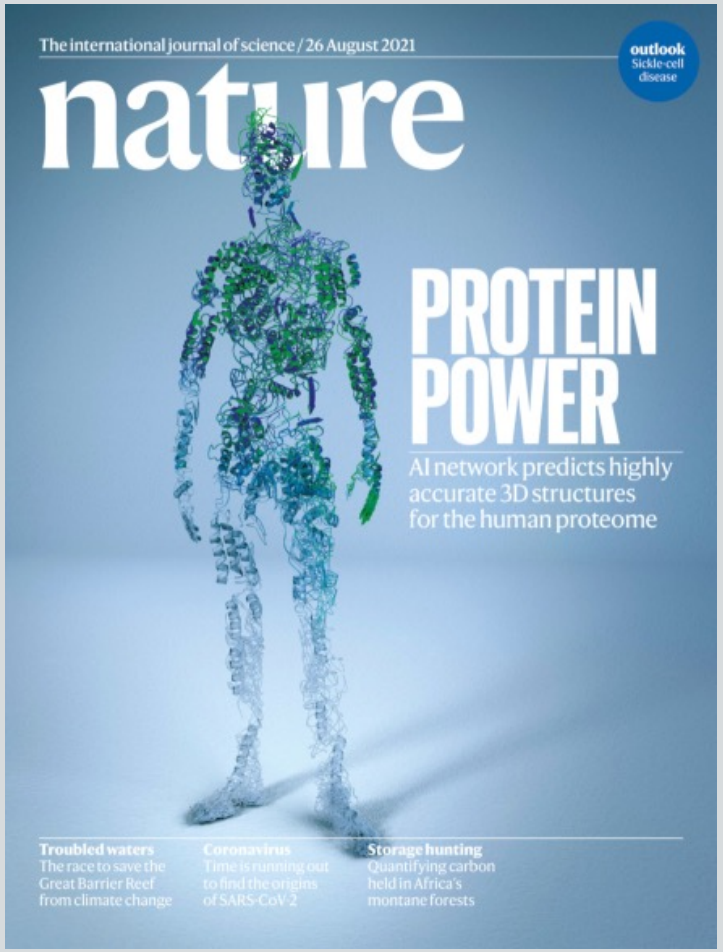
These successes lead to the question of whether it is possible to accurately predict the structure of complexes containing a much wider range of biomolecules, including ligands, ions, nucleic acids and modified residues, within a deep-learning framework. A wide range of predictors for various specific interaction types has been developed[16–28], as well as one generalist method developed concurrently with the present work[29], but the accuracy of such deep-learning attempts has been mixed and often below that of physics-inspired methods[30,31]. Almost all of these methods are also highly specialized to particular interaction types and cannot predict the structure of general biomolecular complexes containing many types of entities.

Here we present AlphaFold 3 (AF3)—a model that is capable of high-accuracy prediction of complexes containing nearly all molecular types present in the Protein Data Bank[32] (PDB) (Fig. 1a,b). In all but one category, it achieves a substantially higher performance than strong methods that specialize in just the given task (Fig. 1c and Extended Data Table 1), including higher accuracy at protein structure and the structure of protein–protein interactions.

This is achieved by a substantial evolution of the AF2 architecture and training procedure (Fig. 1d) both to accommodate more general chemical structures and to improve the data efficiency of learning. The system reduces the amount of multiple-sequence alignment (MSA) processing by replacing the AF2 evoformer with the simpler pairformer module (Fig. 2a). Furthermore it directly predicts the raw atom coordinates with a diffusion module, replacing the AF2 structure module that operated on amino-acid-specific frames and side-chain torsion angles (Fig. 2b). The multiscale nature of the diffusion process (low noise levels induce the network to improve local structure) also enable us to eliminate stereochemical losses and most special handling of bonding patterns in the network, easily accommodating arbitrary chemical components.

[1]Core Contributor, Google DeepMind, London, UK. [2]Core Contributor, Isomorphic Labs, London, UK. [3]Google DeepMind, London, UK. [4]Isomorphic Labs, London, UK. [5]Department of Molecular and Cellular Physiology, Stanford University, Stanford, CA, USA. [6]Department of Computer Science, Princeton University, Princeton, NJ, USA. [7]These authors contributed equally: Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore. [8]These authors jointly supervised this work: Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, John M. Jumper. ✉e-mail: jaderberg@isomorphiclabs.com; dhcontact@google.com; jumper@google.com

Large language models in non-text domains

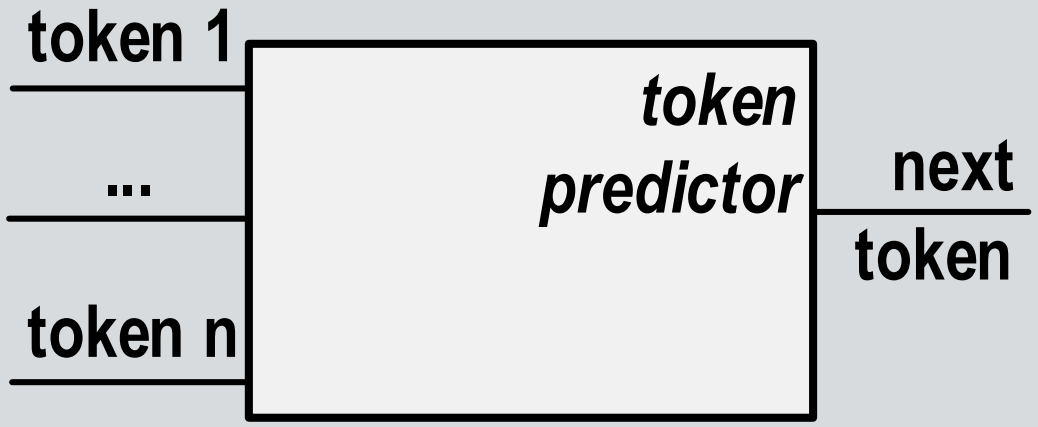


used tokens:

- 20 common amino acid types
- an unknown type
- a gap token
- a mask token

Source:
Jumper et al.: Highly accurate protein structure prediction with AlphaFold. In: Nature, 2021, Vol. 596, pp. 583-589

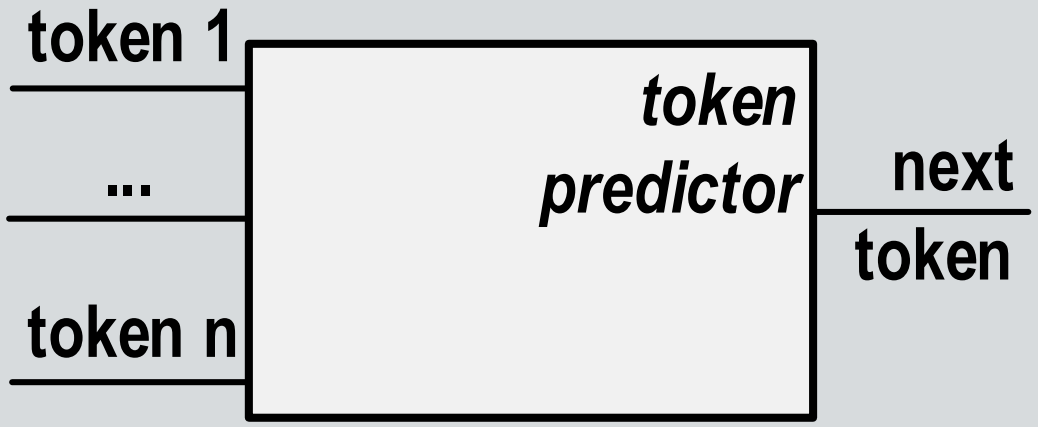
What is a large language model?
A technical view (refined)



What is a token?

- character
- word
- n-gram
- part of a word
- byte pair encoding (BPE)
- amino acid types
- program code
- visual chunks

What is a large language model?
A technical view (refined)



What is a token?

- character
- word
- n-gram
- part of a word
- byte pair encoding (BPE)
- amino acid types

What is a *token* in case of *processes*?

Agenda

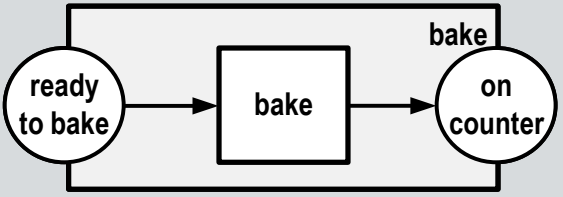
1. Large Language Models
2. Process Models
3. Large Process Models

dfki ai Deutsches Forschungszentrum für Künstliche Intelligenz
German Research Center for Artificial Intelligence

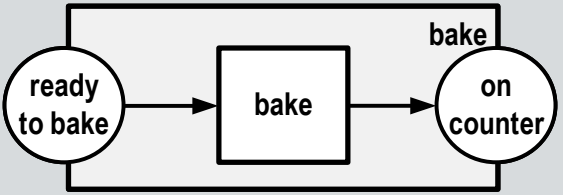
UNIVERSITÄT DES SAARLANDES

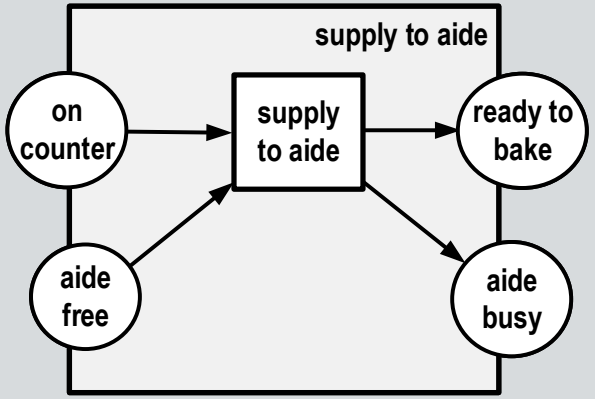
HUMBOLDT-UNIVERSITÄT ZU BERLIN

What is a process model?
A small bakery process

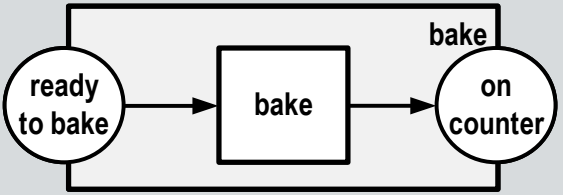


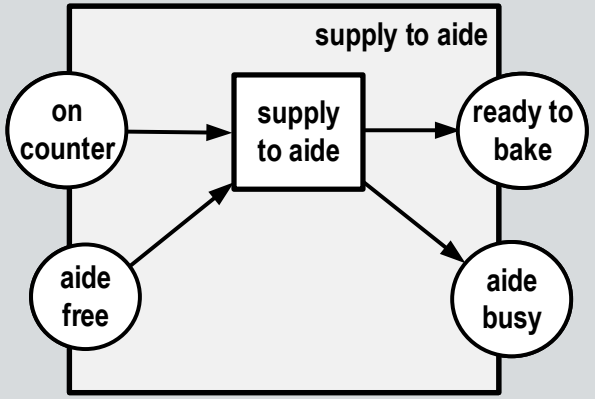
What is a process model?
A small bakery process

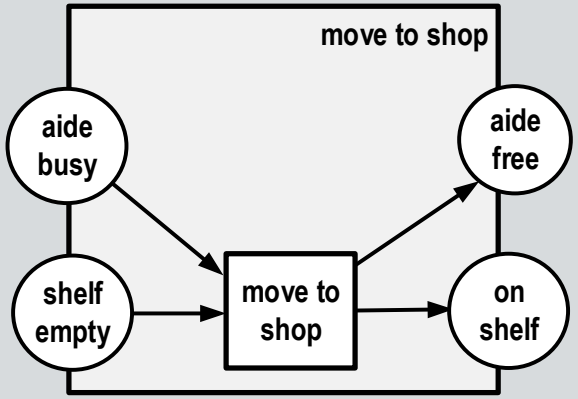




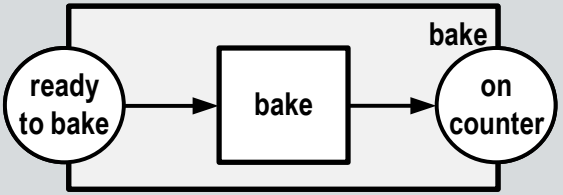
What is a process model?
A small bakery process

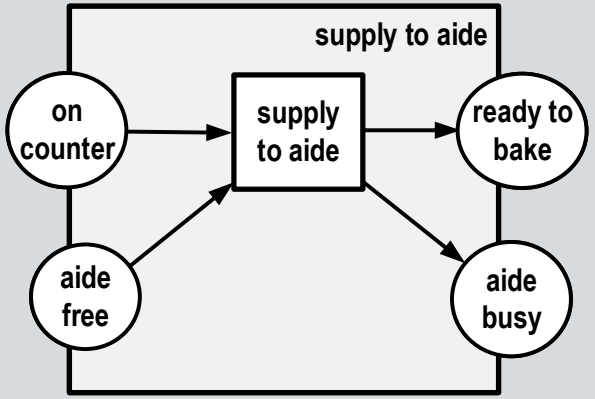


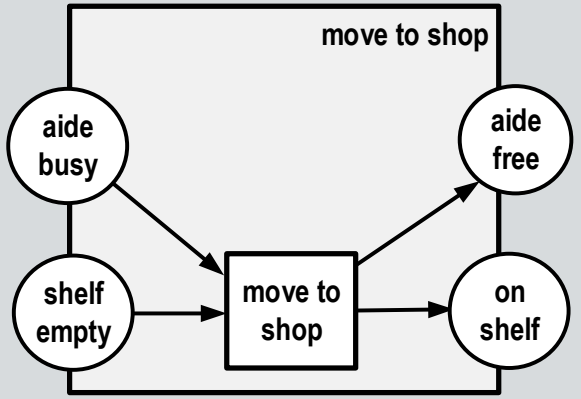


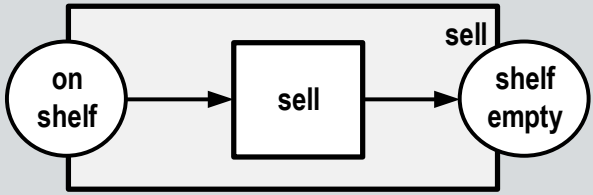


What is a process model?
A small bakery process

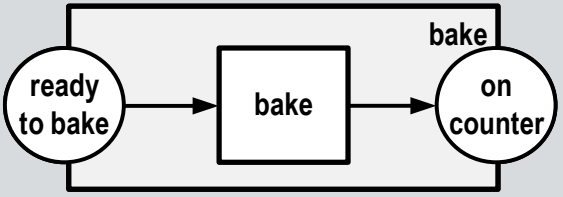




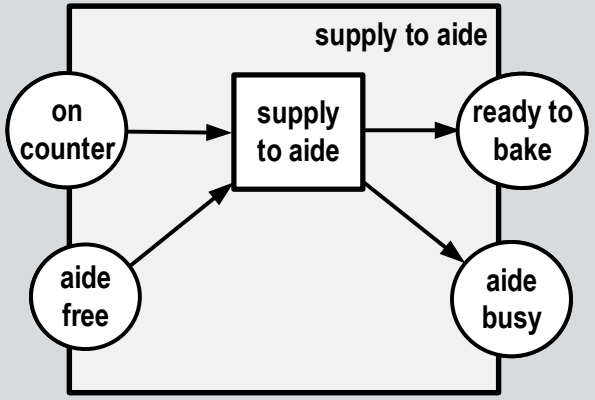


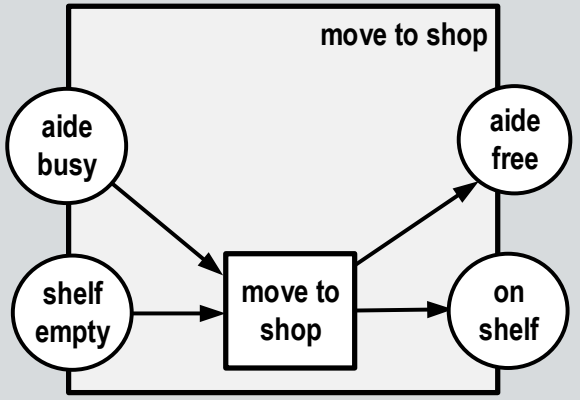


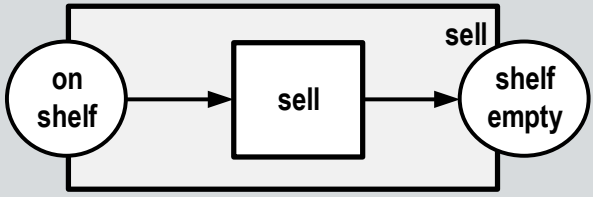
What is a process model?
A small bakery process



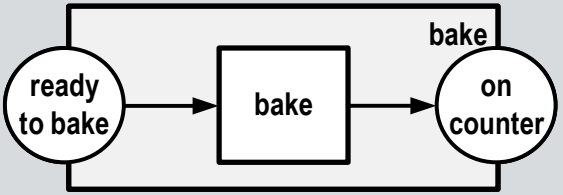
four steps describe behavior

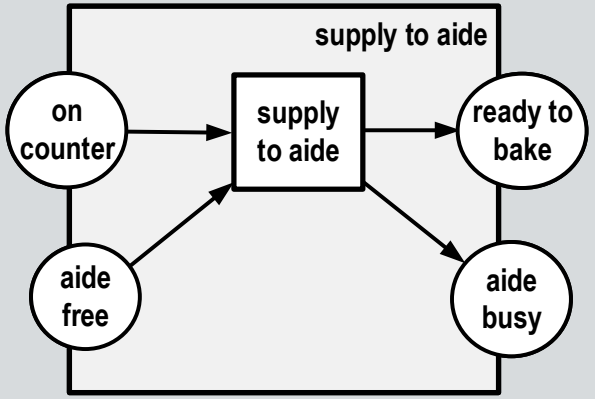


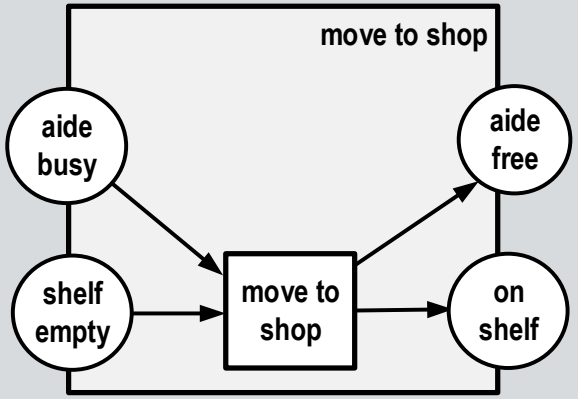


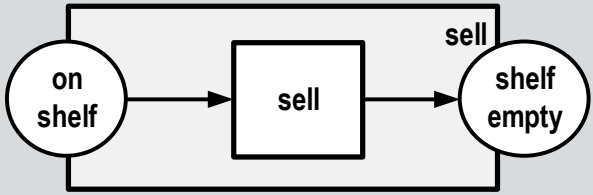


What is a process model?
A small bakery process







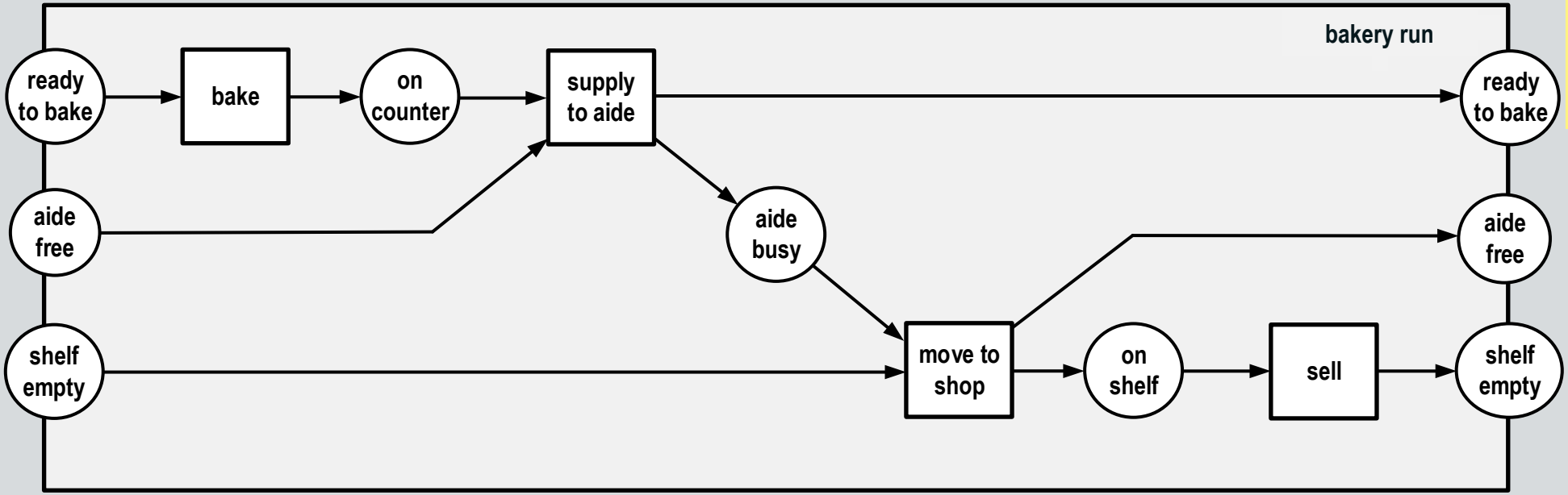


four steps
describe behavior

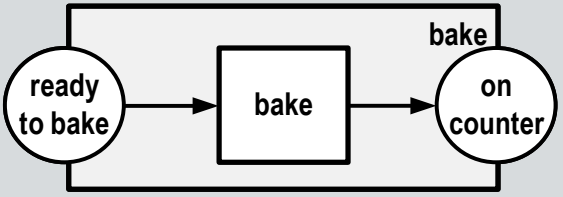
composed steps

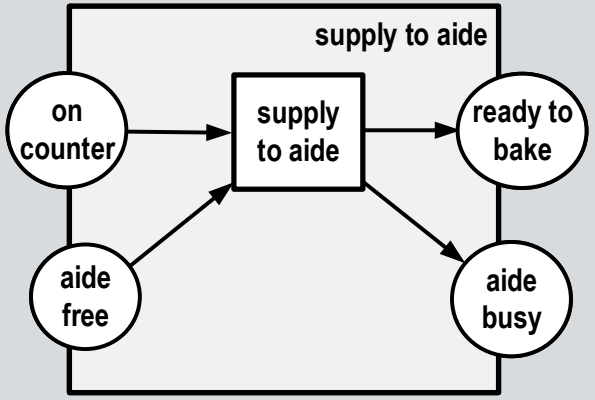
bakery run:

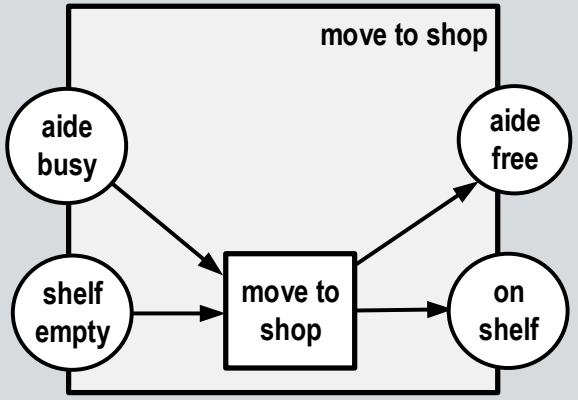
bake • supply to aide • move to shop • sell

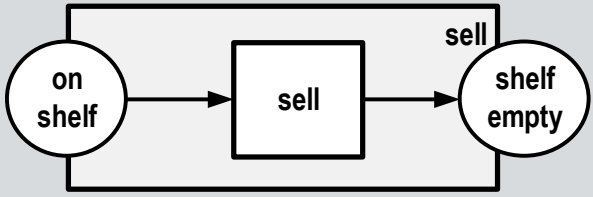


A little bit larger

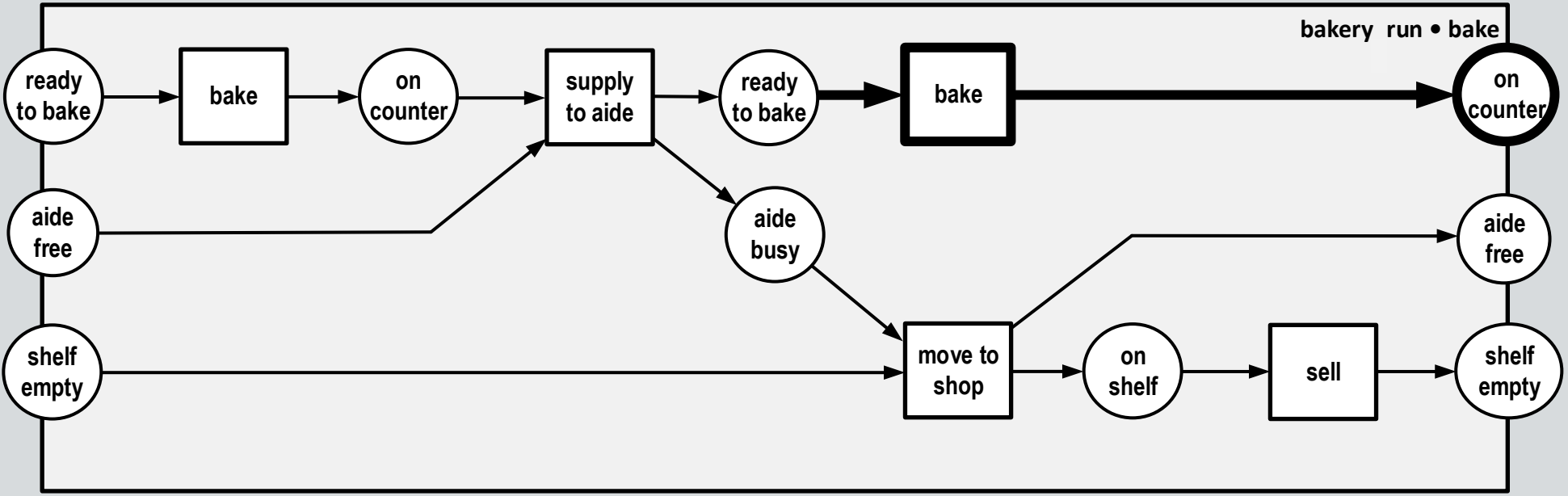








four steps
describe behavior



composed steps

bakery run:

bake • supply to aide • move to shop • sell

extension

bakery run • bake

no more

totally ordered!

Larger bakery processes

bakery run $=_{def}$ bake • supply to aide • move to shop • sell

Larger bakery processes

bakery run $=_{def}$ bake • supply to aide • move to shop • sell

(1)-bakery runs $=_{def}$ bakery run

Larger bakery processes

bakery run $=_{def}$ bake • supply to aide • move to shop • sell

(1)-bakery runs $=_{def}$ bakery run

(2)-bakery runs $=_{def}$ (1)-bakery runs • bakery run

Larger bakery processes

bakery run $=_{def}$ bake • supply to aide • move to shop • sell

(1)-bakery runs $=_{def}$ bakery run

(2)-bakery runs $=_{def}$ (1)-bakery runs • bakery run

(3)-bakery runs $=_{def}$ (2)-bakery runs • bakery run

Larger bakery processes

bakery run $=_{def}$ bake • supply to aide • move to shop • sell

(1)-bakery runs $=_{def}$ bakery run

(2)-bakery runs $=_{def}$ (1)-bakery runs • bakery run

(3)-bakery runs $=_{def}$ (2)-bakery runs • bakery run

…

(n+1)-bakery runs $=_{def}$ (n)-bakery runs • bakery run

Larger bakery processes

bakery run $=_{def}$ bake • supply to aide • move to shop • sell

(1)-bakery runs $=_{def}$ bakery run

(2)-bakery runs $=_{def}$ (1)-bakery runs • bakery run

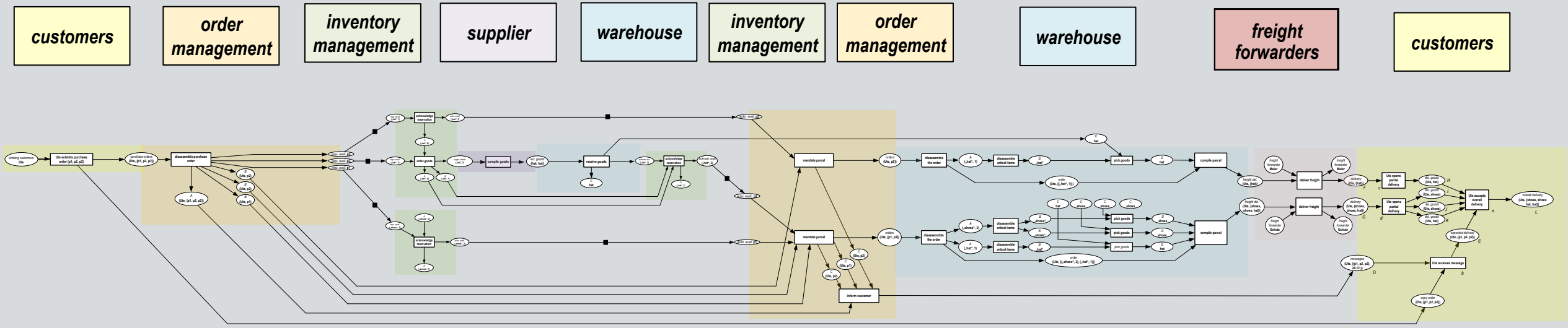
(3)-bakery runs $=_{def}$ (2)-bakery runs • bakery run

…

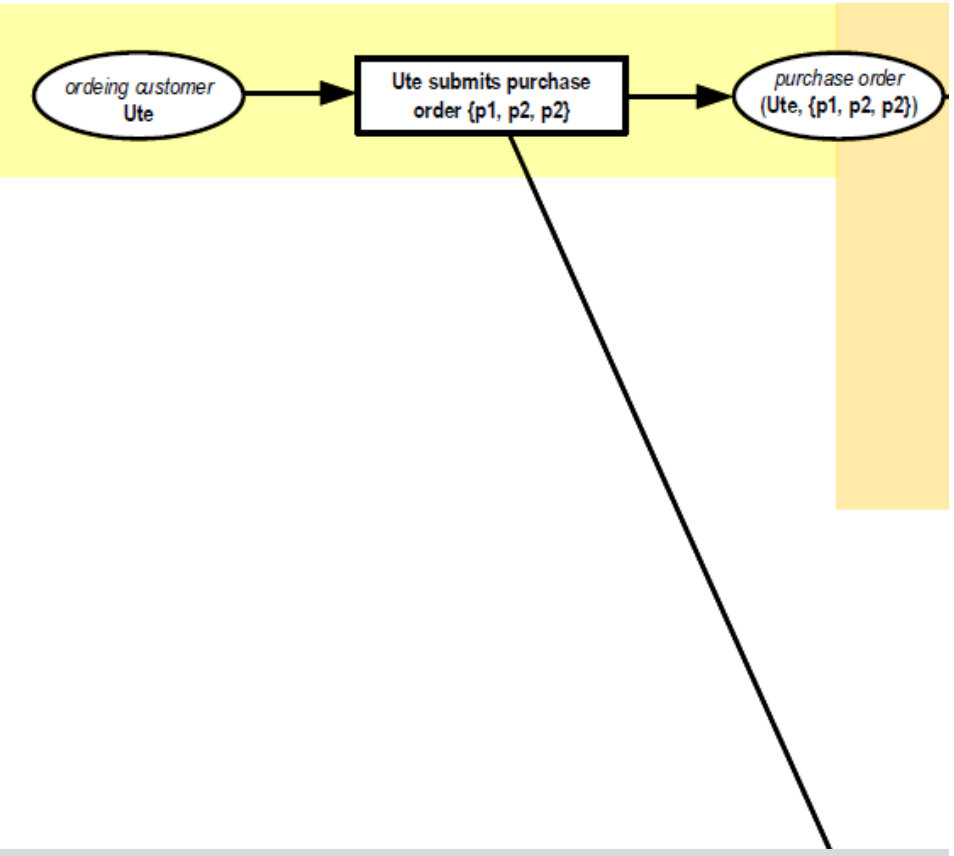
(n+1)-bakery runs $=_{def}$ (n)-bakery runs • bakery run

This is an infinite process – that’s really large…

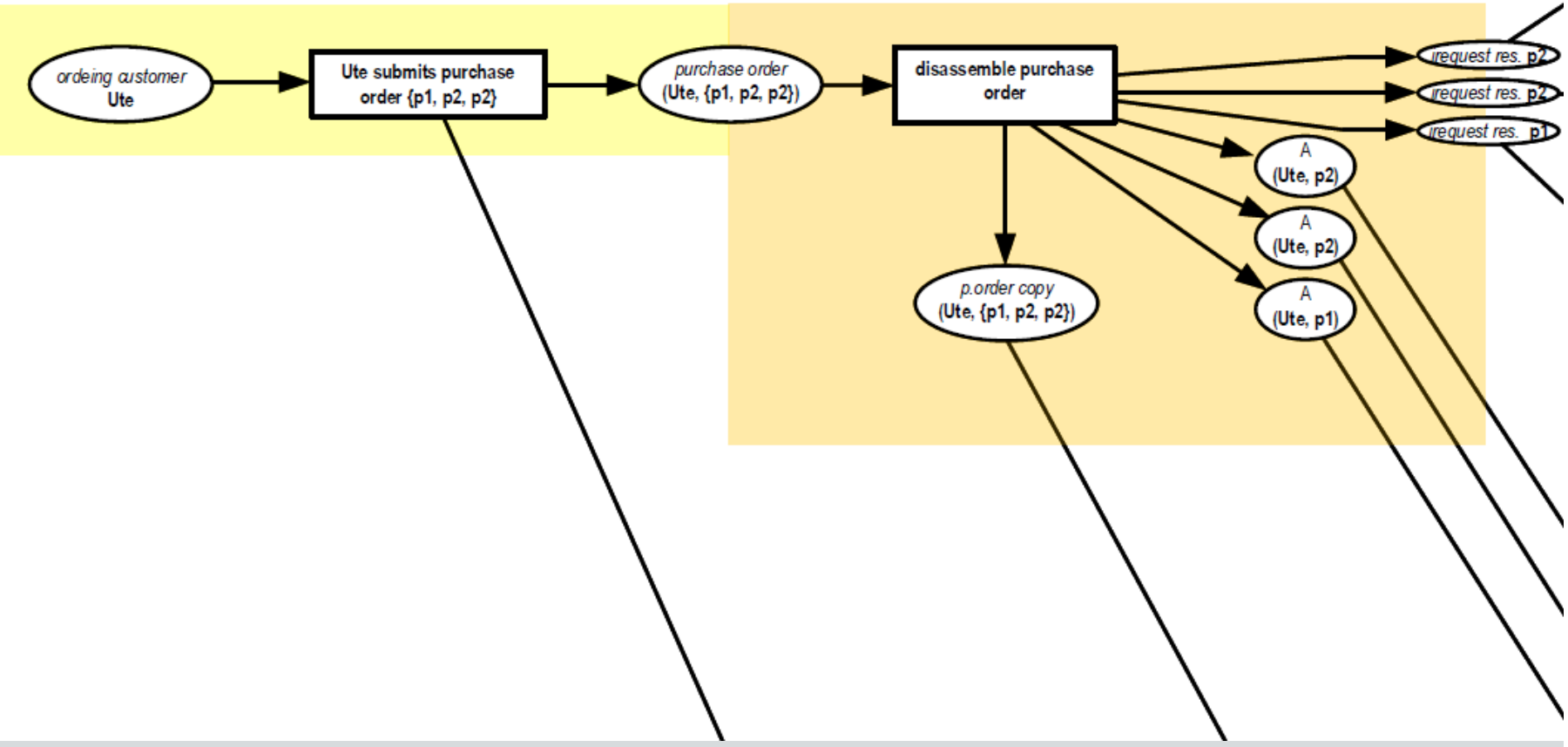
A more realistic large process: order fulfillment



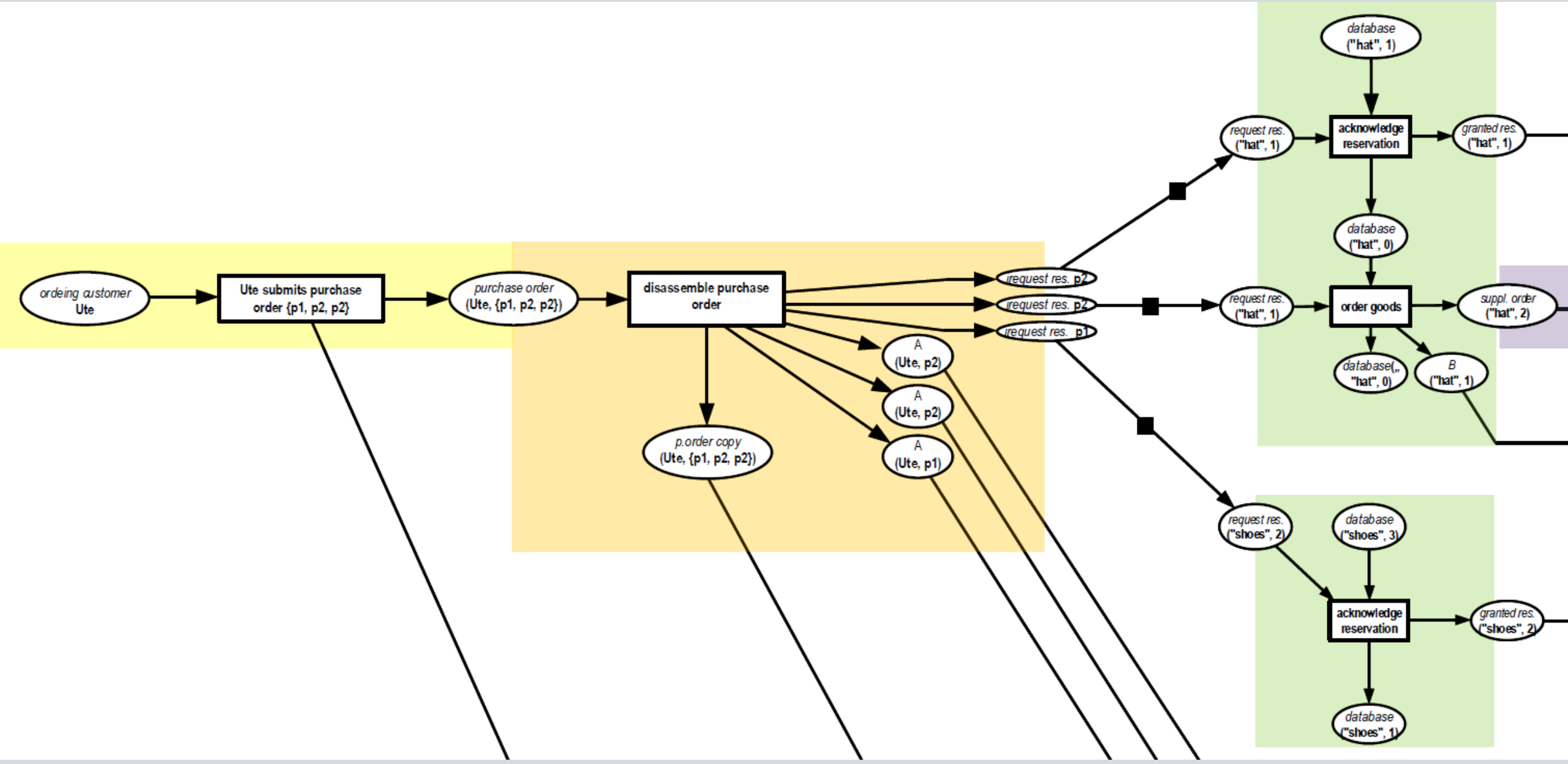
Some details



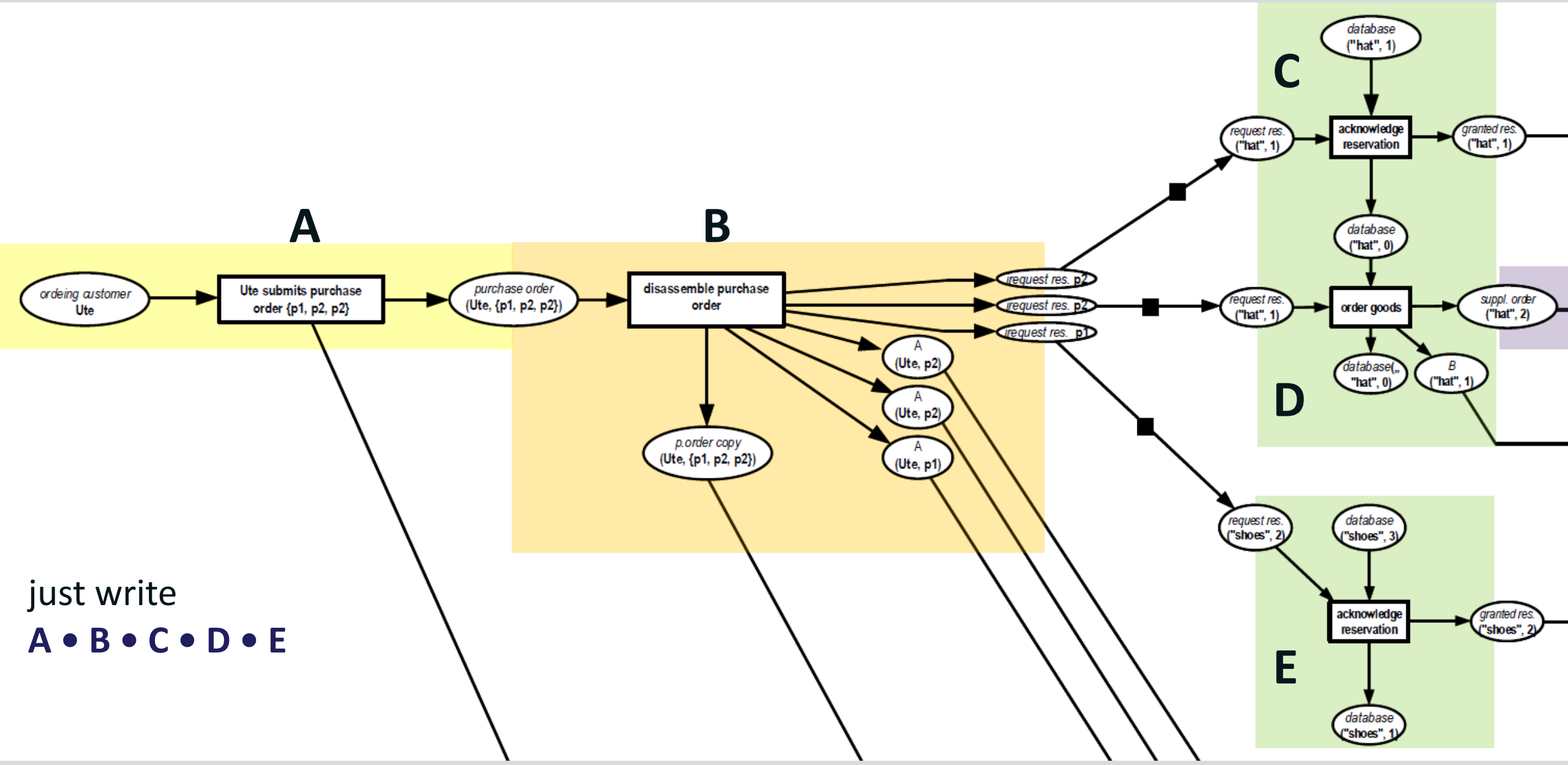
Some details



Some details



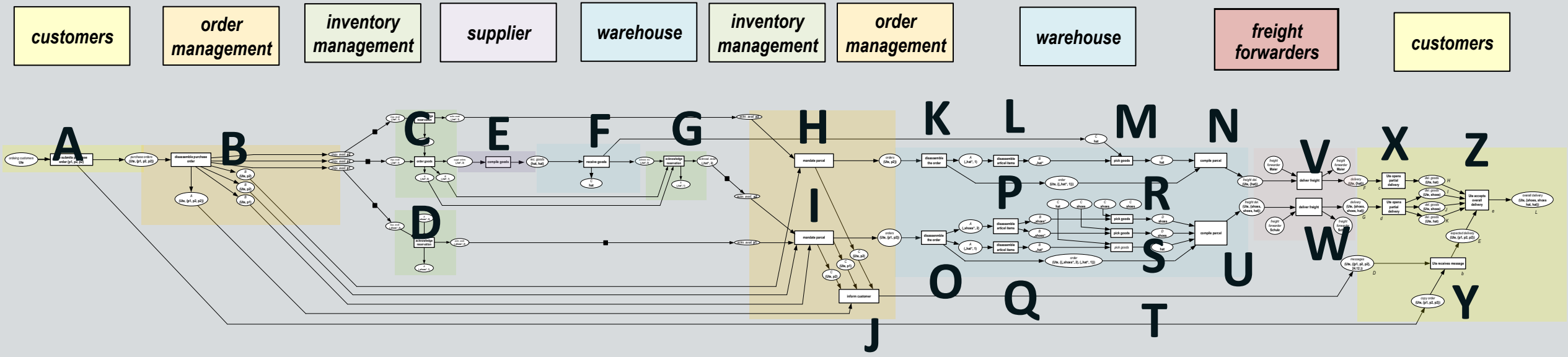
Some details



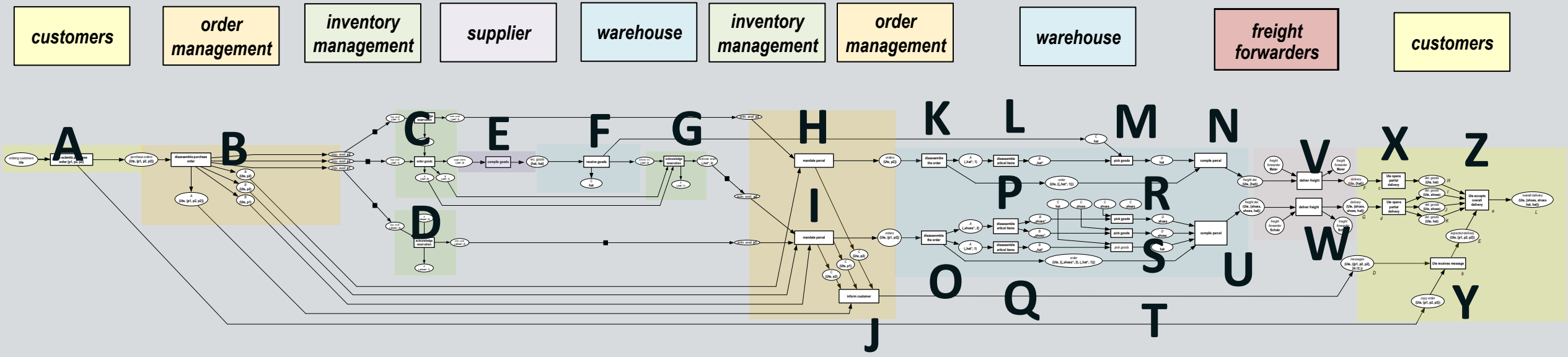
just write

A • B • C • D • E

A more realistic large process: order fulfillment



A more realistic large process: order fulfillment



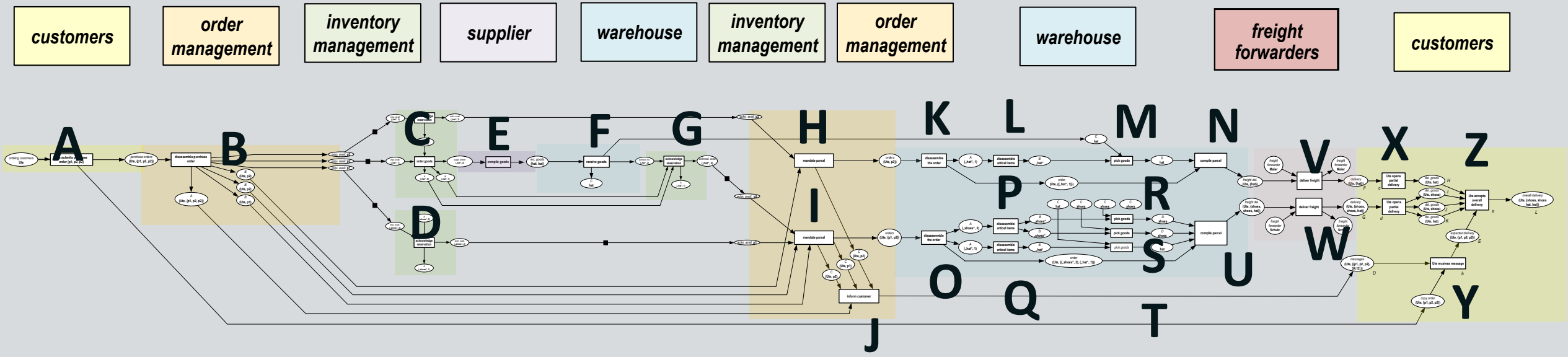
just write

A • B • C • D • E • F • G • H • I • J • K • L • M • N • O • P • Q • R • S • T • U • V • W • X • Y • Z

three challenges

1. objects and data
2. interaction between cases
3. composition

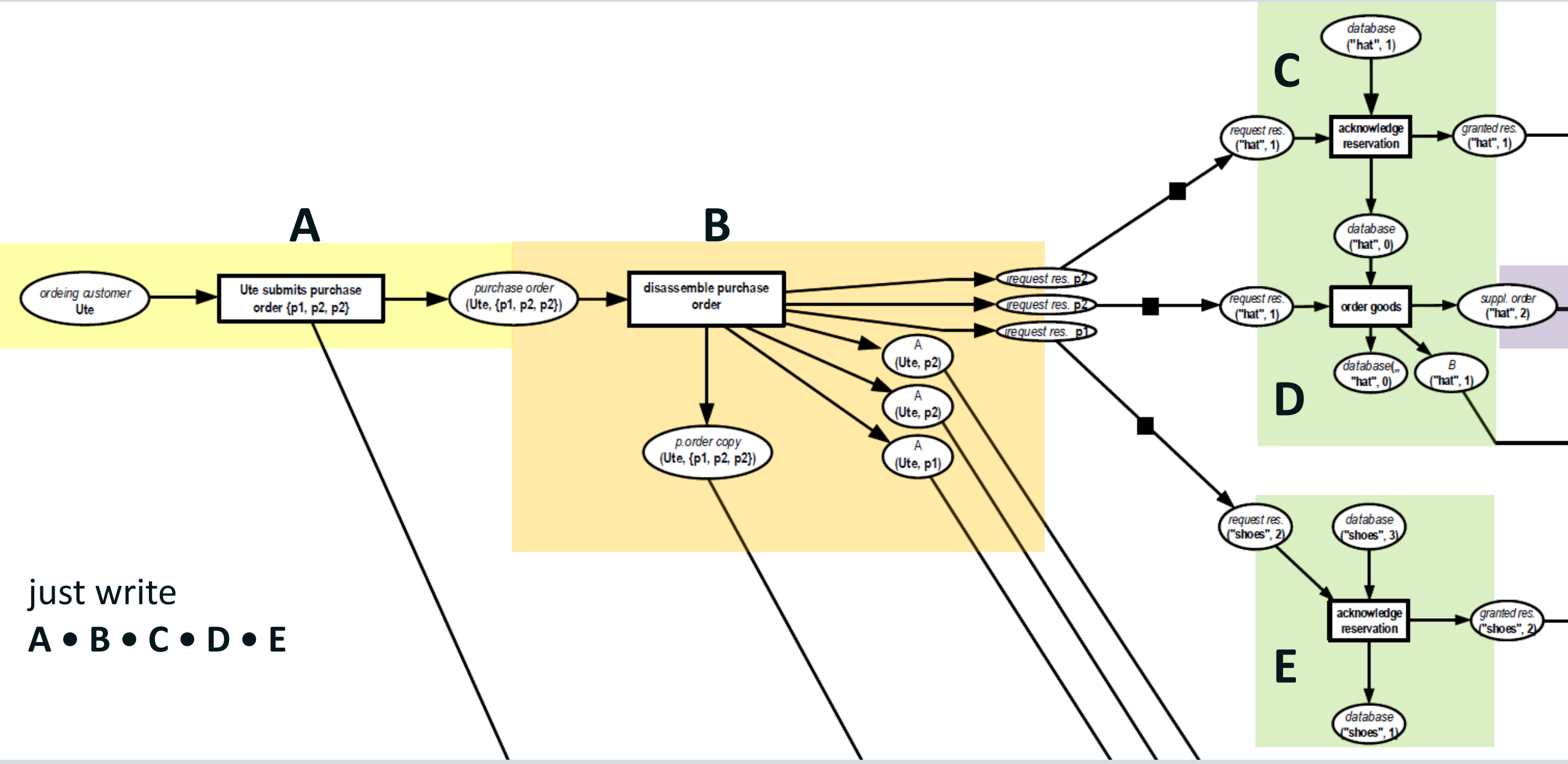
c large process: order fulfillment



just write

A • B • C • D • E • F • G • H • I • J • K • L • M • N • O • P • Q • R • S • T • U • V • W • X • Y • Z

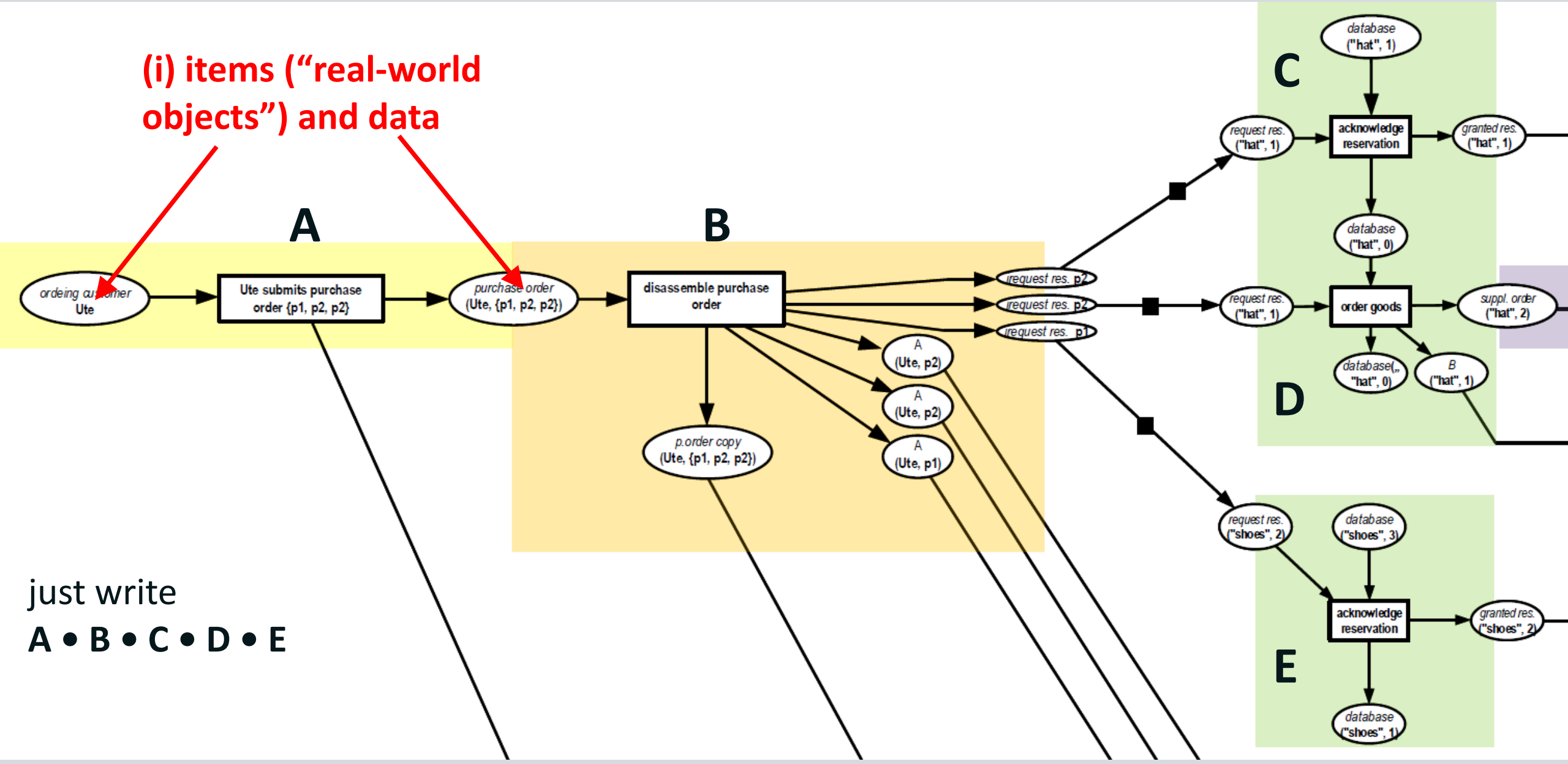
The composition calculus deals with the three challenges



just write

A • B • C • D • E

The composition calculus deals with the three challenges



just write

A • B • C • D • E

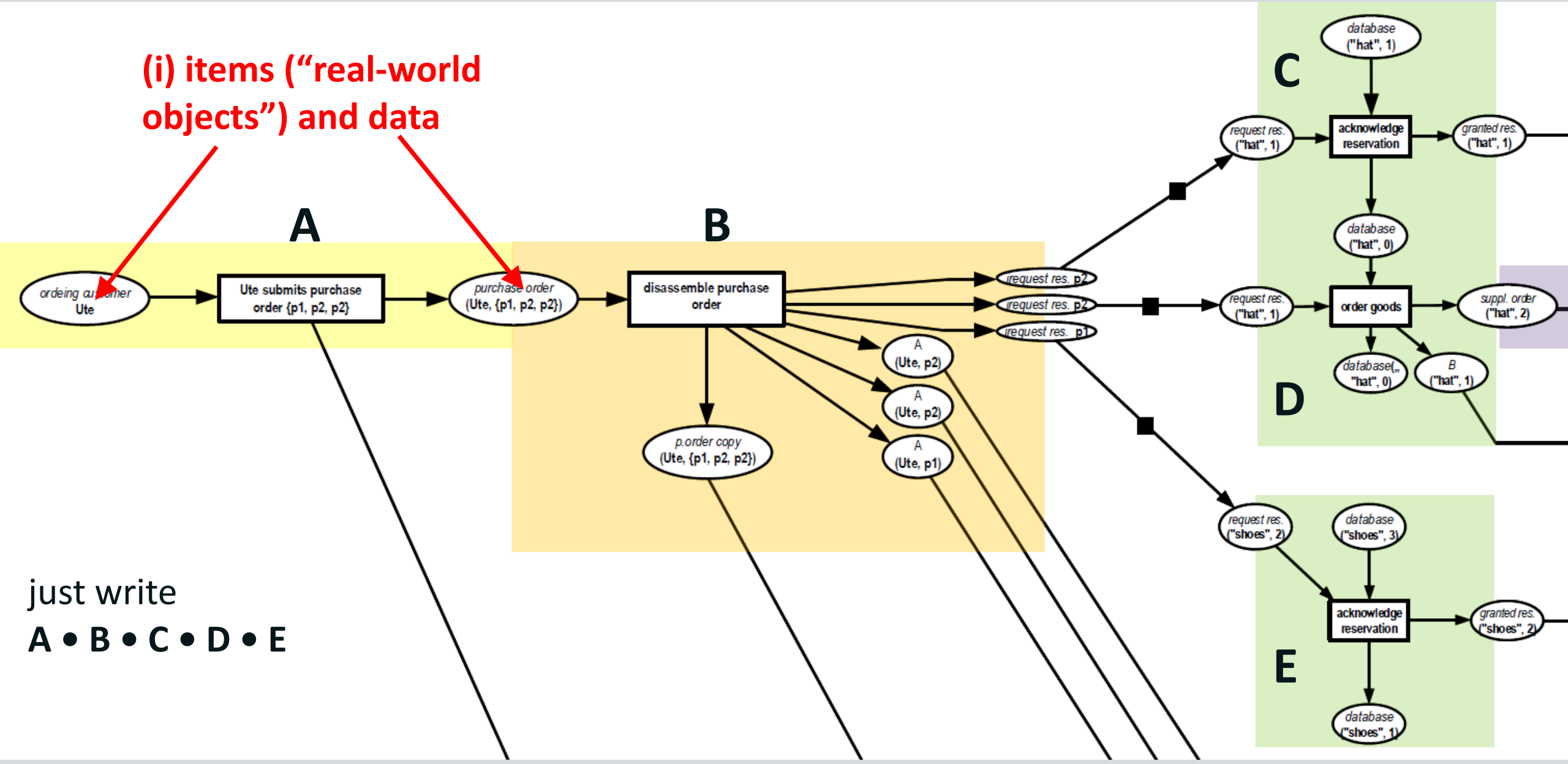
Excursus: Why is it important to distinguish between items (“real-world objects”) and data?

a fish head (“item”)

price of a fish head
(“data”)

Wolfgang can eat the *fish head*.
But he cannot eat the *price* of the fish head!

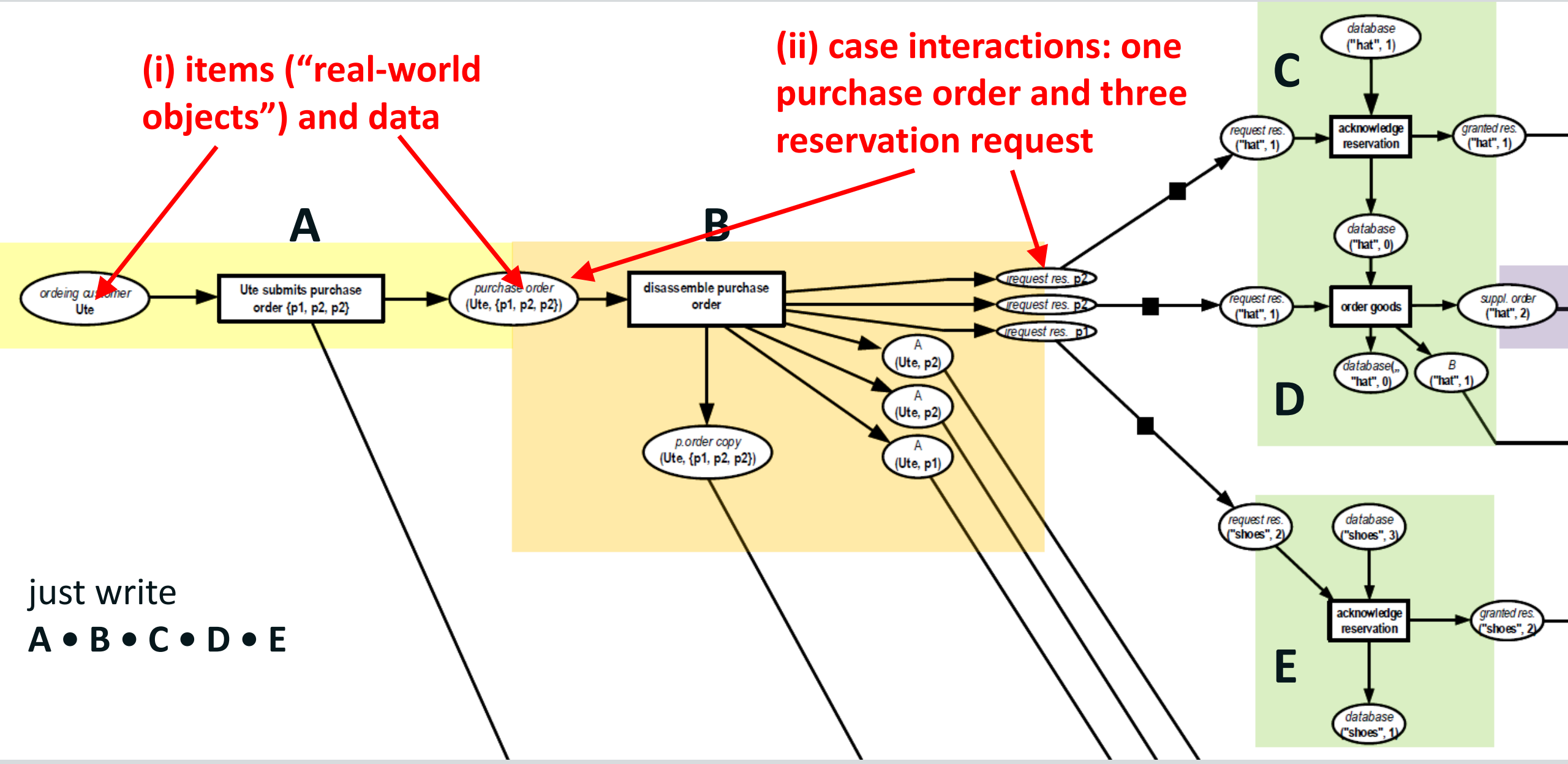
The composition calculus deals with the three challenges



just write

A • B • C • D • E

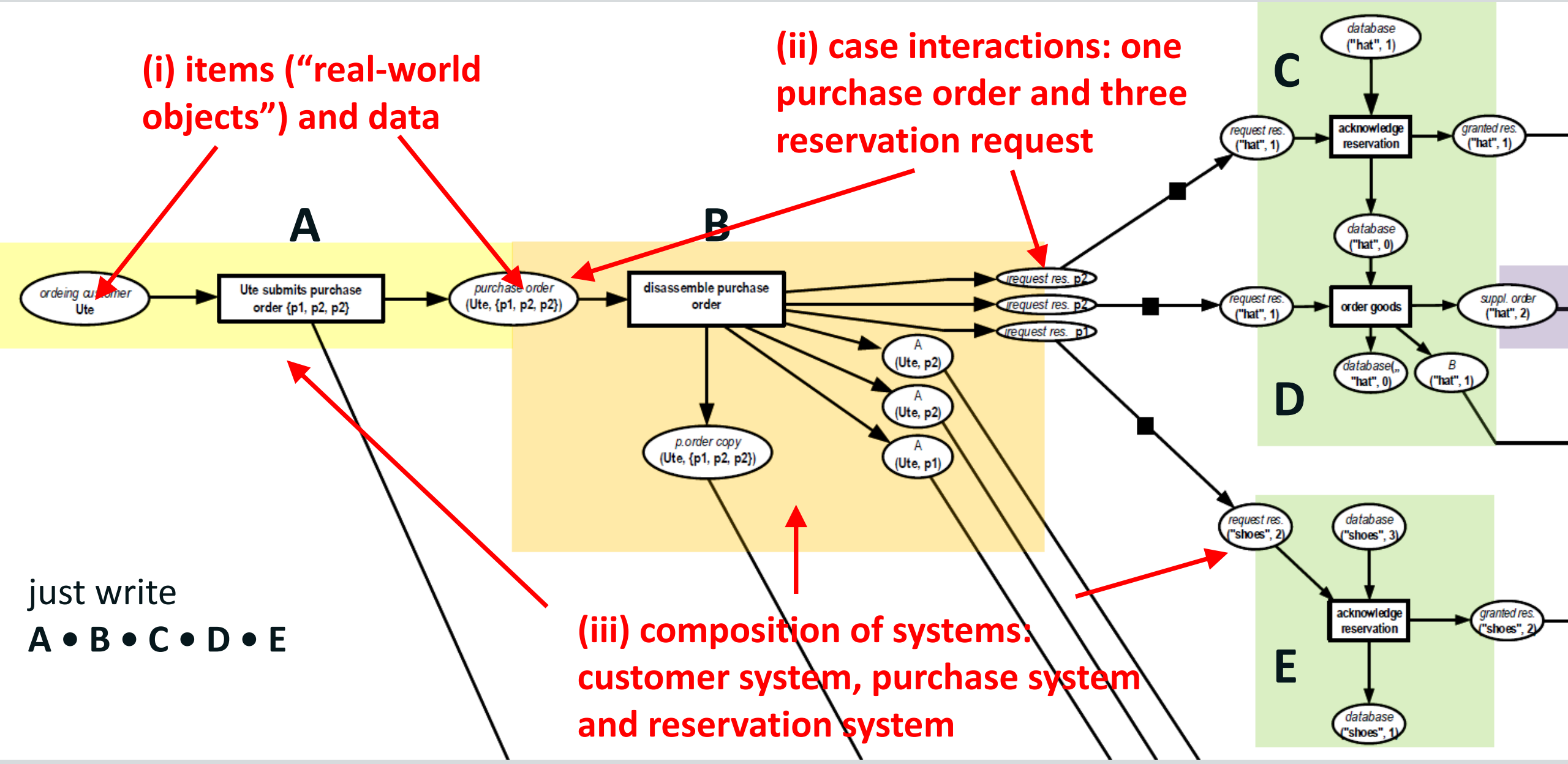
The composition calculus deals with the three challenges



just write

A • B • C • D • E

The composition calculus deals with the three challenges



Agenda

1. Large Language Models
2. Process Models
3. Large Process Models

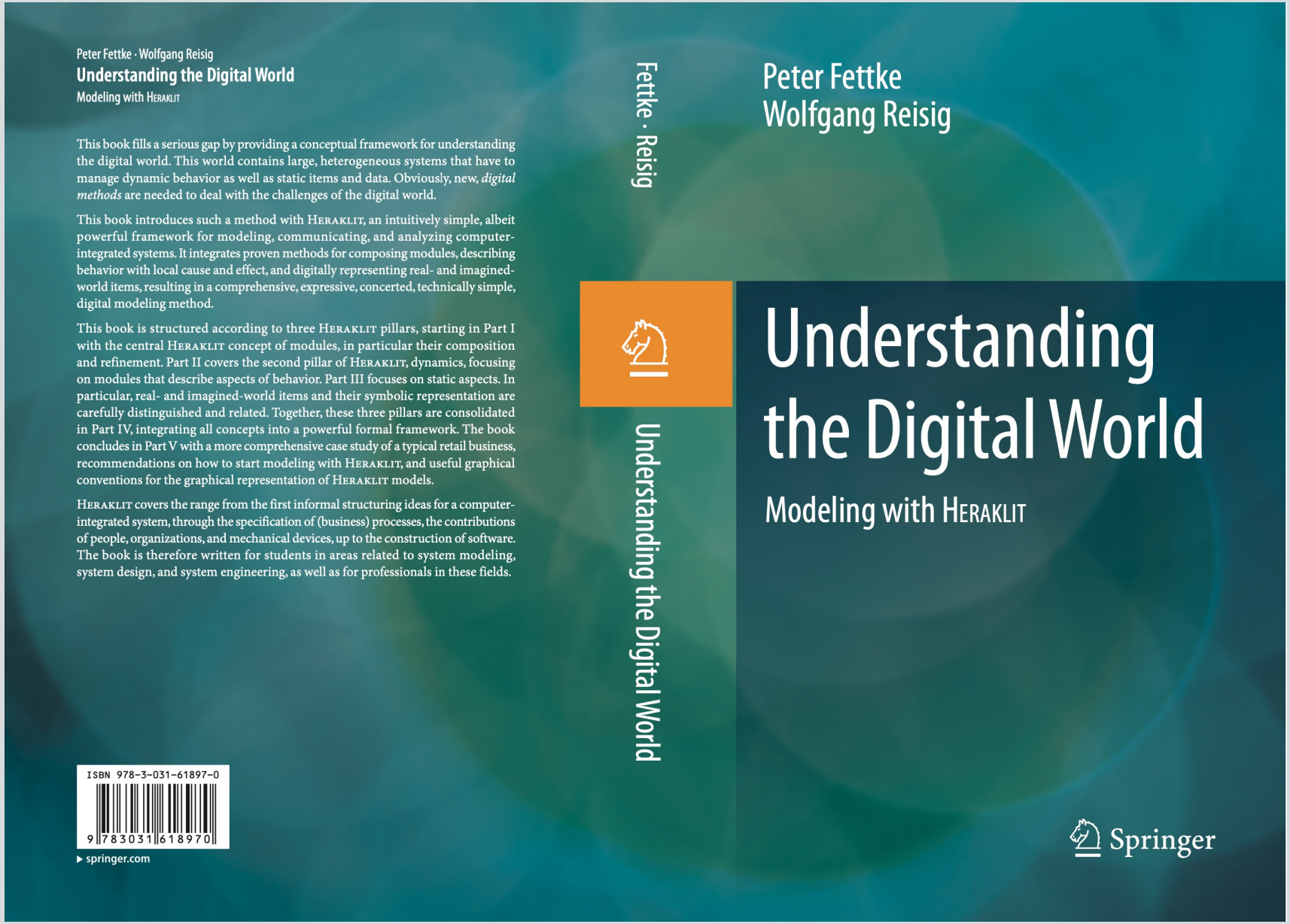
dfki ai Deutsches Forschungszentrum für Künstliche Intelligenz
German Research Center for Artificial Intelligence

UNIVERSITÄT DES SAARLANDES

HUMBOLDT-UNIVERSITÄT ZU BERLIN

Commercial break

Commercial break



Idea for a large process model

idea of a large language model is applied in the context of business process management

Idea for a large process model

idea of a large language model is applied in the context of business process management

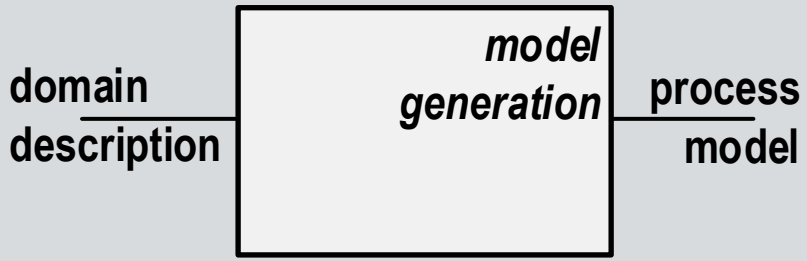
three approaches:

(1) fined-tuned large language model

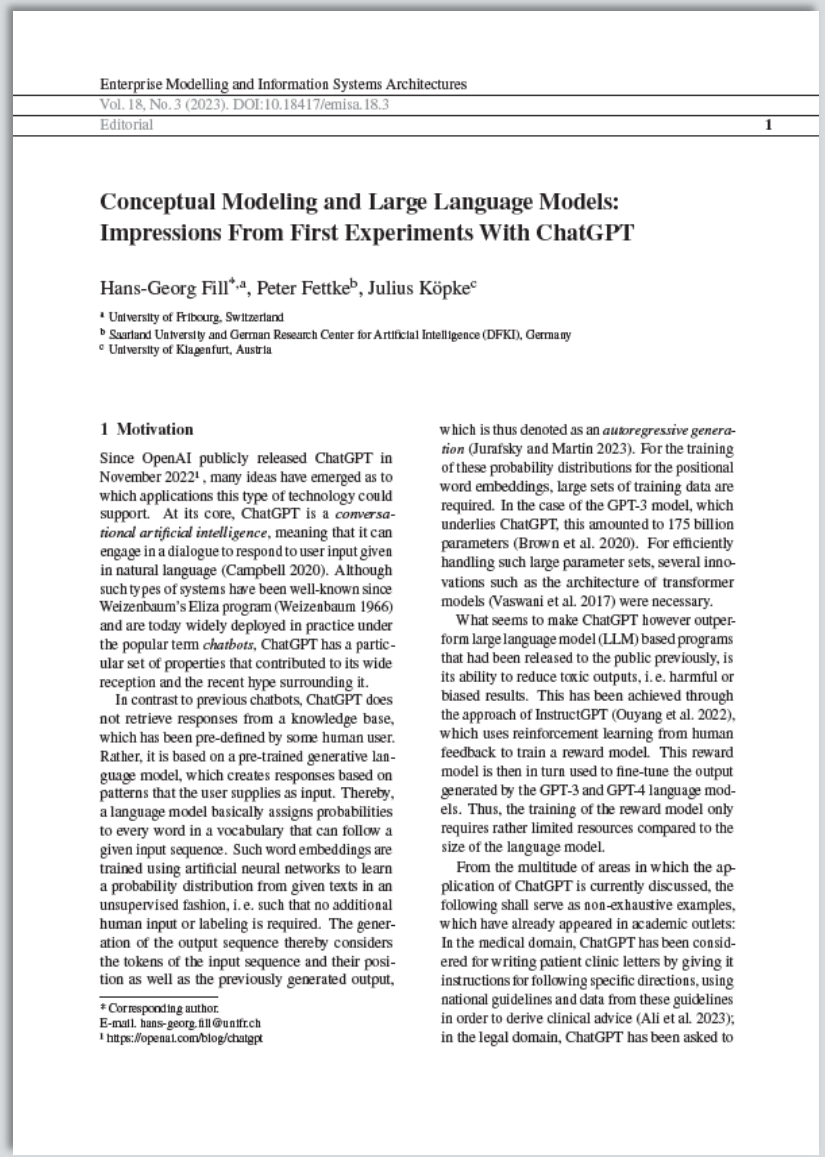
(2) large language model agent

(3) native large language model

(1) Large process model as a fine-tuned large language model



(1) Large process model as a fine-tuned large language model: example



Enterprise Modelling and Information Systems Architectures
Vol. 18, No. 3 (2023). DOI:10.18417/emisa.18.3

Conceptual Modeling and Large Language Models: Impressions From First Experiments With ChatGPT

Hans-Georg Fill*,[a], Peter Fettke[b], Julius Köpke[c]

[a] University of Fribourg, Switzerland
[b] Saarland University and German Research Center for Artificial Intelligence (DFKI), Germany
[c] University of Klagenfurt, Austria

1 Motivation

Since OpenAI publicly released ChatGPT in November 2022[1], many ideas have emerged as to which applications this type of technology could support. At its core, ChatGPT is a *conversational artificial intelligence*, meaning that it can engage in a dialogue to respond to user input given in natural language (Campbell 2020). Although such types of systems have been well-known since Weizenbaum's Eliza program (Weizenbaum 1966) and are today widely deployed in practice under the popular term *chatbots*, ChatGPT has a particular set of properties that contributed to its wide reception and the recent hype surrounding it.

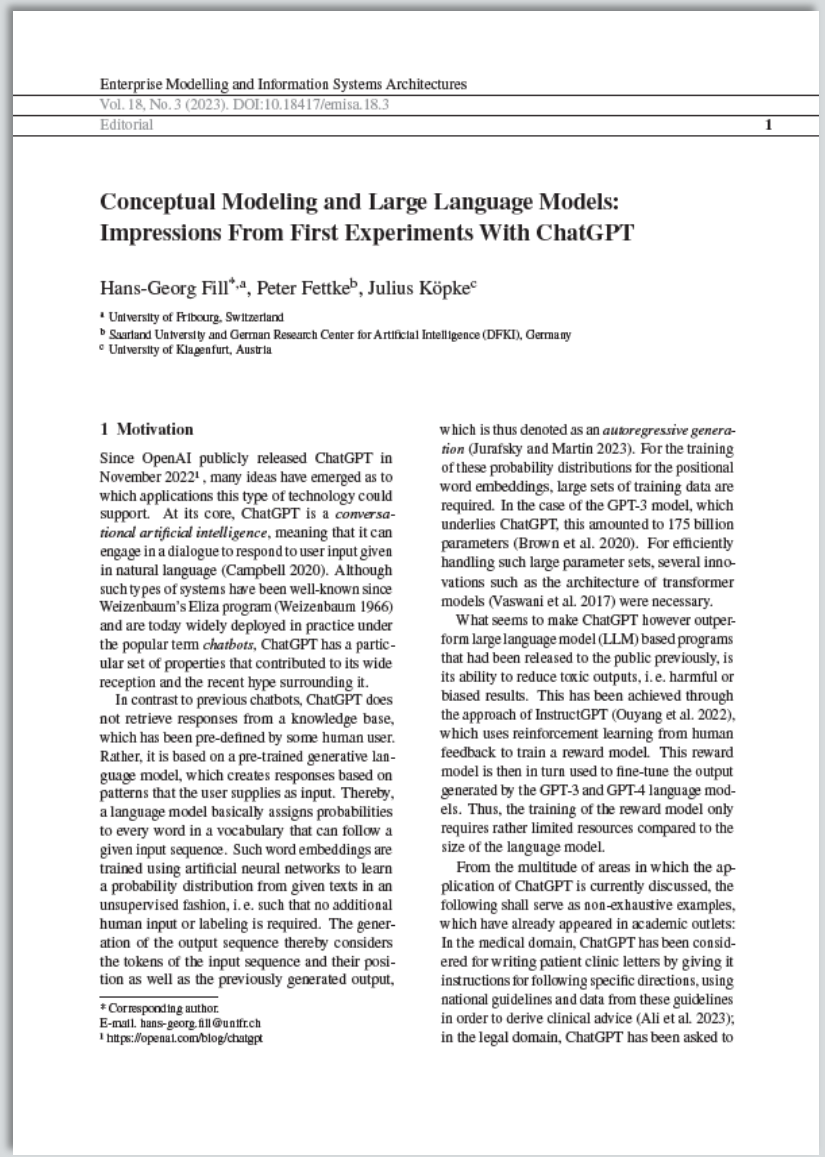
In contrast to previous chatbots, ChatGPT does not retrieve responses from a knowledge base, which has been pre-defined by some human user. Rather, it is based on a pre-trained generative language model, which creates responses based on patterns that the user supplies as input. Thereby, a language model basically assigns probabilities to every word in a vocabulary that can follow a given input sequence. Such word embeddings are trained using artificial neural networks to learn a probability distribution from given texts in an unsupervised fashion, i.e. such that no additional human input or labeling is required. The generation of the output sequence thereby considers the tokens of the input sequence and their position as well as the previously generated output, which is thus denoted as an *autoregressive generation* (Jurafsky and Martin 2023). For the training of these probability distributions for the positional word embeddings, large sets of training data are required. In the case of the GPT-3 model, which underlies ChatGPT, this amounted to 175 billion parameters (Brown et al. 2020). For efficiently handling such large parameter sets, several innovations such as the architecture of transformer models (Vaswani et al. 2017) were necessary.

What seems to make ChatGPT however outperform large language model (LLM) based programs that had been released to the public previously, is its ability to reduce toxic outputs, i.e. harmful or biased results. This has been achieved through the approach of InstructGPT (Ouyang et al. 2022), which uses reinforcement learning from human feedback to train a reward model. This reward model is then in turn used to fine-tune the output generated by the GPT-3 and GPT-4 language models. Thus, the training of the reward model only requires rather limited resources compared to the size of the language model.

From the multitude of areas in which the application of ChatGPT is currently discussed, the following shall serve as non-exhaustive examples, which have already appeared in academic outlets: In the medical domain, ChatGPT has been considered for writing patient clinic letters by giving it instructions for following specific directions, using national guidelines and data from these guidelines in order to derive clinical advice (Ali et al. 2023); in the legal domain, ChatGPT has been asked to

* Corresponding author.
E-mail. hans-georg.fill@unifr.ch
[1] https://openai.com/blog/chatgpt

(1) Large process model as a fine-tuned large language model: example



Enterprise Modelling and Information Systems Architectures
Vol. 18, No. 3 (2023). DOI:10.18417/emisa.18.3

Conceptual Modeling and Large Language Models: Impressions From First Experiments With ChatGPT

Hans-Georg Fill*,[a], Peter Fettke[b], Julius Köpke[c]

[a] University of Fribourg, Switzerland
[b] Saarland University and German Research Center for Artificial Intelligence (DFKI), Germany
[c] University of Klagenfurt, Austria

1 Motivation

Since OpenAI publicly released ChatGPT in November 2022[1], many ideas have emerged as to which applications this type of technology could support. At its core, ChatGPT is a *conversational artificial intelligence*, meaning that it can engage in a dialogue to respond to user input given in natural language (Campbell 2020). Although such types of systems have been well-known since Weizenbaum's Eliza program (Weizenbaum 1966) and are today widely deployed in practice under the popular term *chatbots*, ChatGPT has a particular set of properties that contributed to its wide reception and the recent hype surrounding it.

In contrast to previous chatbots, ChatGPT does not retrieve responses from a knowledge base, which has been pre-defined by some human user. Rather, it is based on a pre-trained generative language model, which creates responses based on patterns that the user supplies as input. Thereby, a language model basically assigns probabilities to every word in a vocabulary that can follow a given input sequence. Such word embeddings are trained using artificial neural networks to learn a probability distribution from given texts in an unsupervised fashion, i.e. such that no additional human input or labeling is required. The generation of the output sequence thereby considers the tokens of the input sequence and their position as well as the previously generated output, which is thus denoted as an *autoregressive generation* (Jurafsky and Martin 2023). For the training of these probability distributions for the positional word embeddings, large sets of training data are required. In the case of the GPT-3 model, which underlies ChatGPT, this amounted to 175 billion parameters (Brown et al. 2020). For efficiently handling such large parameter sets, several innovations such as the architecture of transformer models (Vaswani et al. 2017) were necessary.

What seems to make ChatGPT however outperform large language model (LLM) based programs that had been released to the public previously, is its ability to reduce toxic outputs, i.e. harmful or biased results. This has been achieved through the approach of InstructGPT (Ouyang et al. 2022), which uses reinforcement learning from human feedback to train a reward model. This reward model is then in turn used to fine-tune the output generated by the GPT-3 and GPT-4 language models. Thus, the training of the reward model only requires rather limited resources compared to the size of the language model.

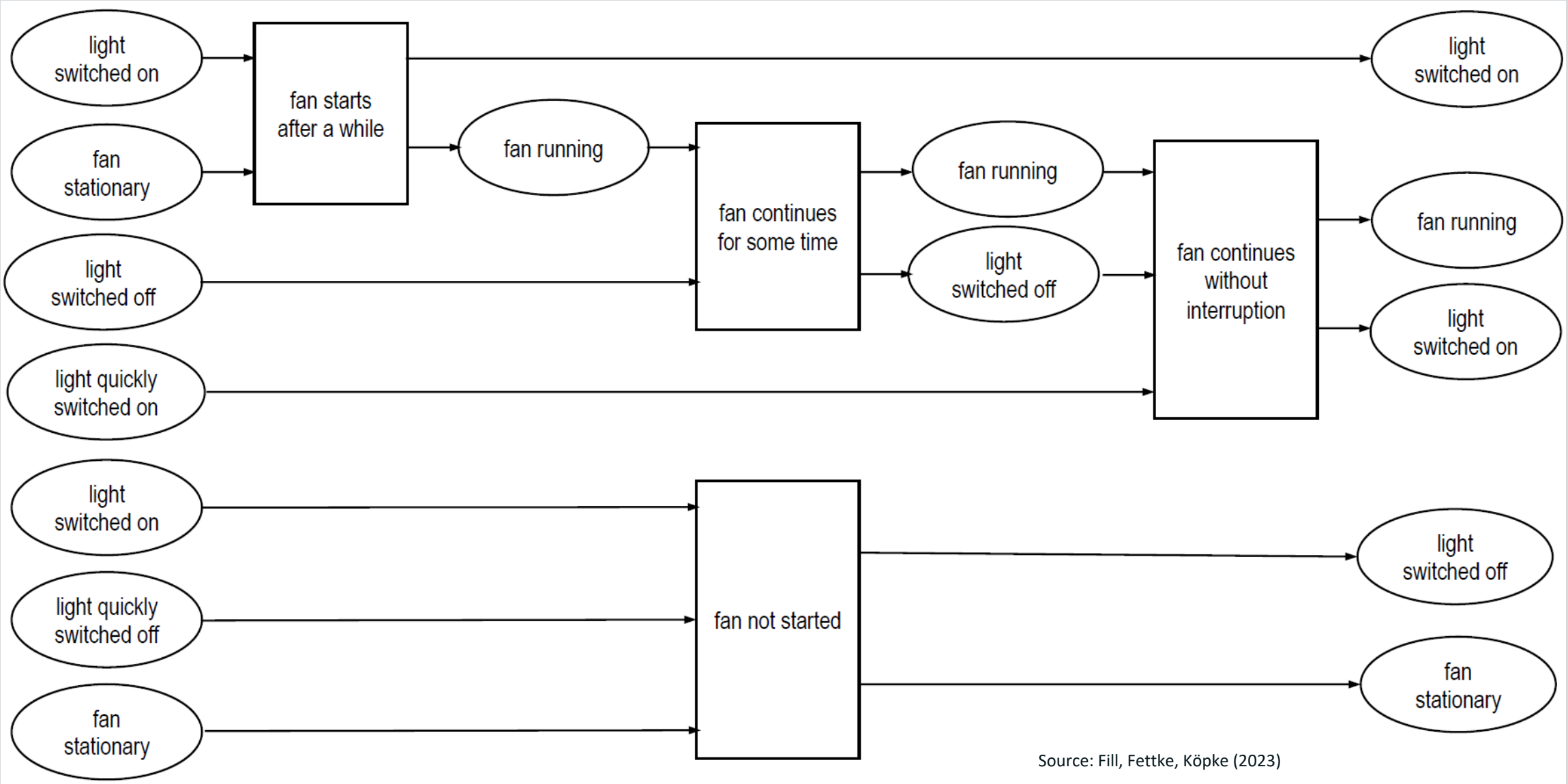
From the multitude of areas in which the application of ChatGPT is currently discussed, the following shall serve as non-exhaustive examples, which have already appeared in academic outlets: In the medical domain, ChatGPT has been considered for writing patient clinic letters by giving it instructions for following specific directions, using national guidelines and data from these guidelines in order to derive clinical advice (Ali et al. 2023); in the legal domain, ChatGPT has been asked to

* Corresponding author.
E-mail: hans-georg.fill@unifr.ch
[1] https://openai.com/blog/chatgpt

Task Definition 5

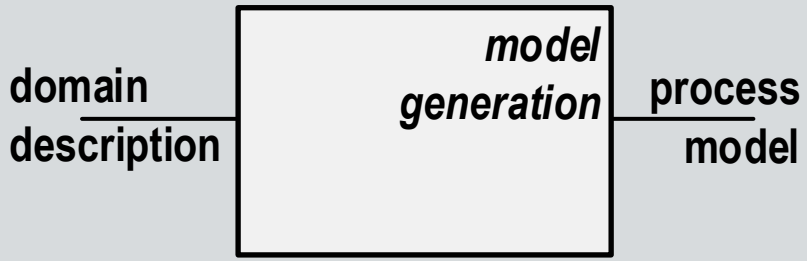
Lighting and ventilation of a bathroom: If the light is switched on when the fan is stationary, the fan also starts after a while. Then, when the light is turned off, the fan continues to run for some time. If the light is turned on first and then turned off quickly when the fan is stationary, the fan will not start at all. If the light is first switched off and then quickly switched on again when the fan is running, the fan continues to run without interruption.

(1) Large process model as a fine-tuned large language model: example

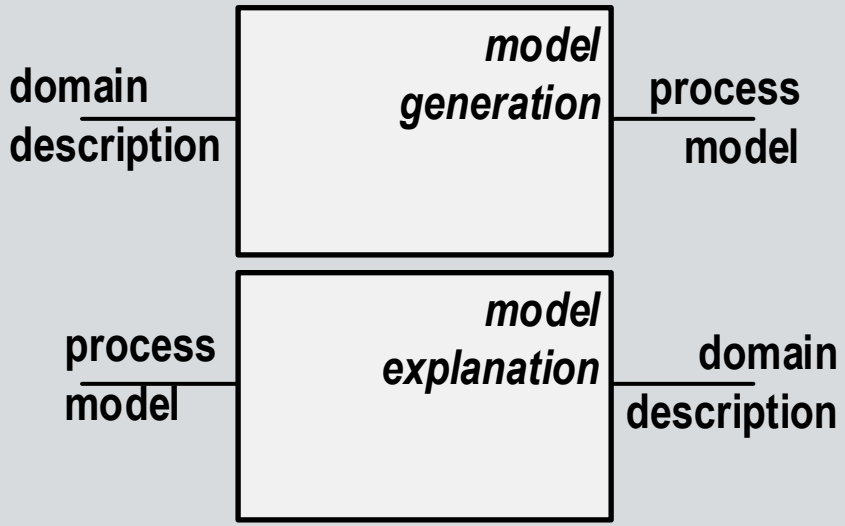


Source: Fill, Fettke, Köpke (2023)

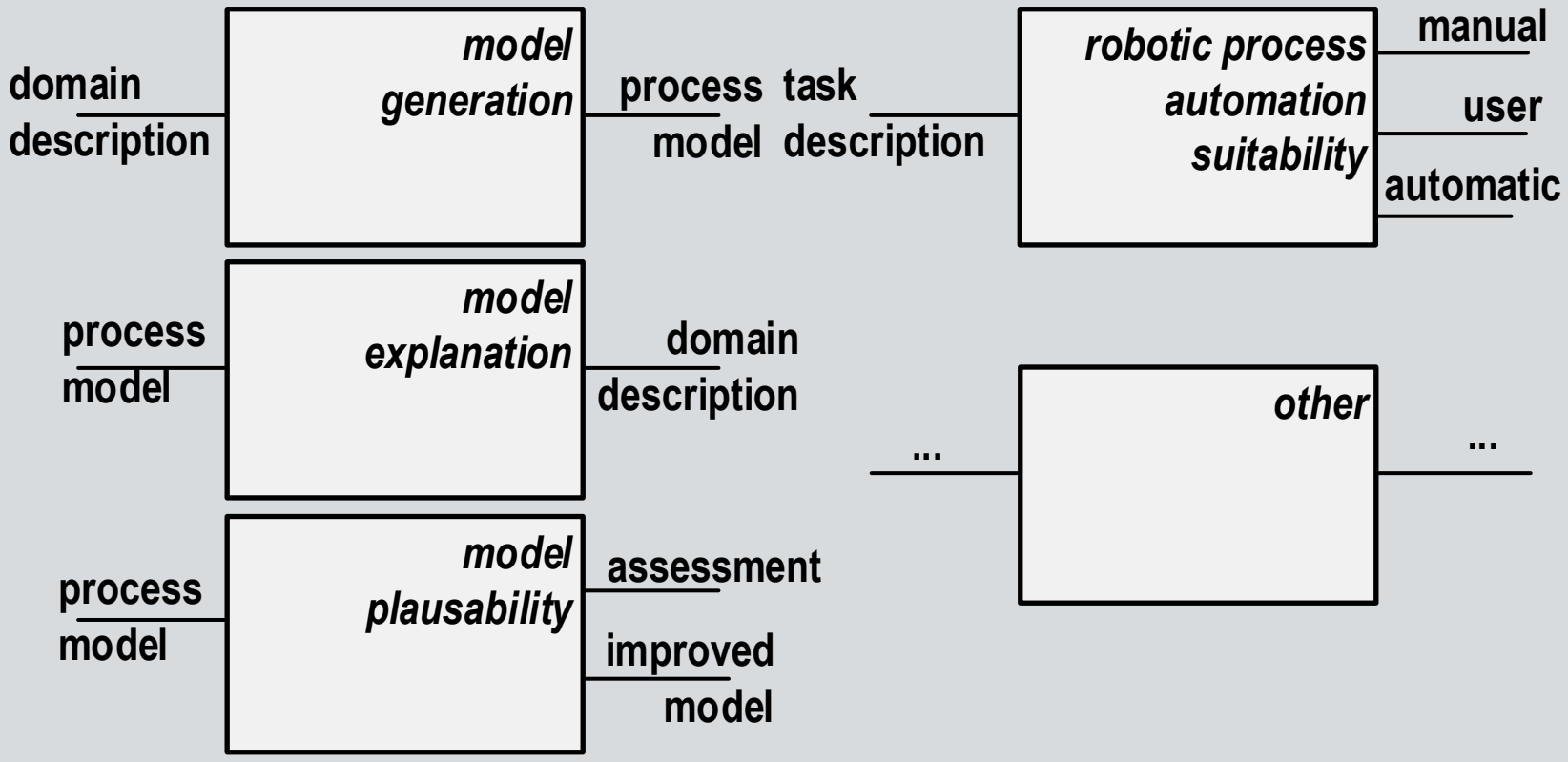
(1) Large process model as a fine-tuned large language model



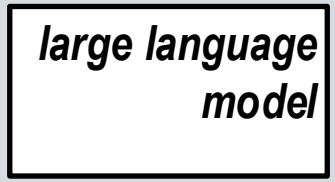
(1) Large process model as a fine-tuned large language model



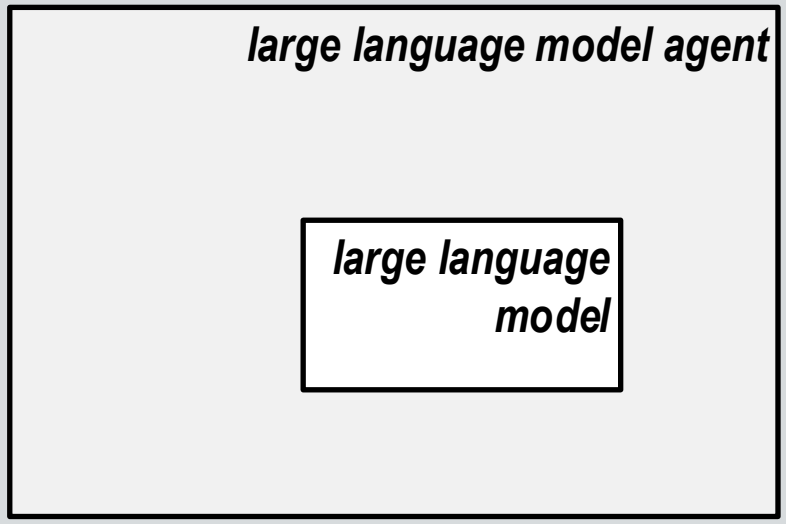
(1) Large process model as a fine-tuned large language model



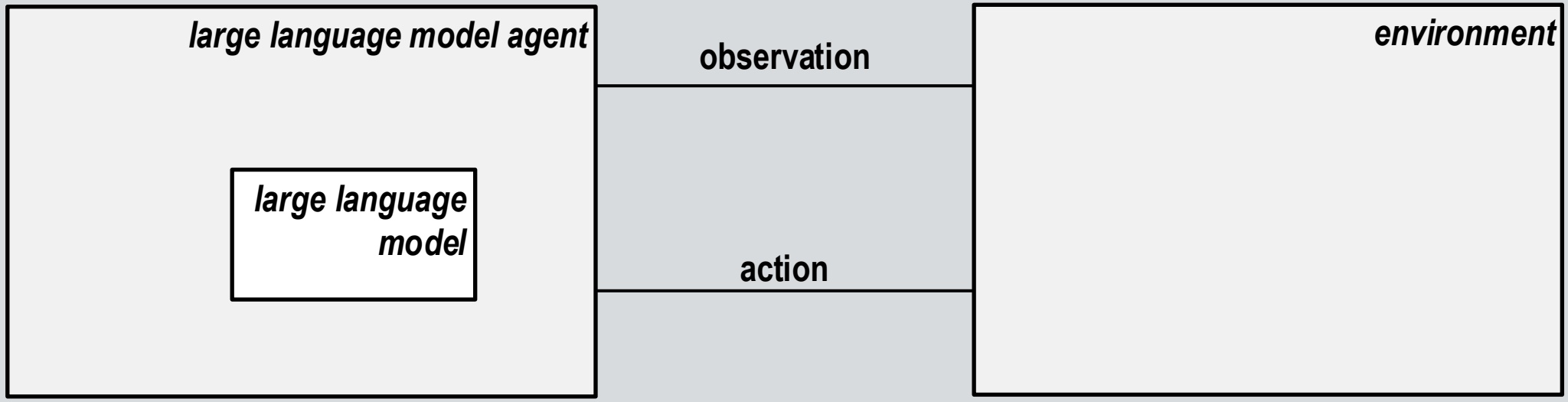
(2) Large process model as a large process model agent



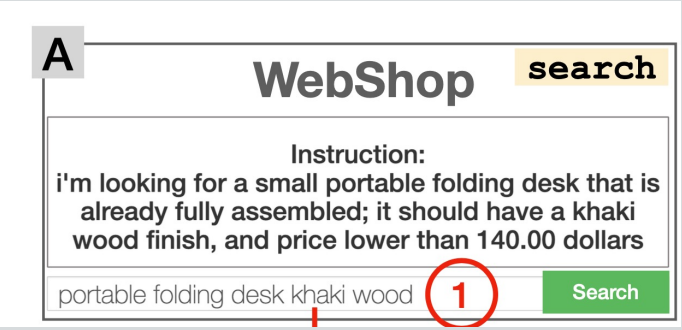
(2) Large process model as a large process model agent



(2) Large process model as a large process model agent

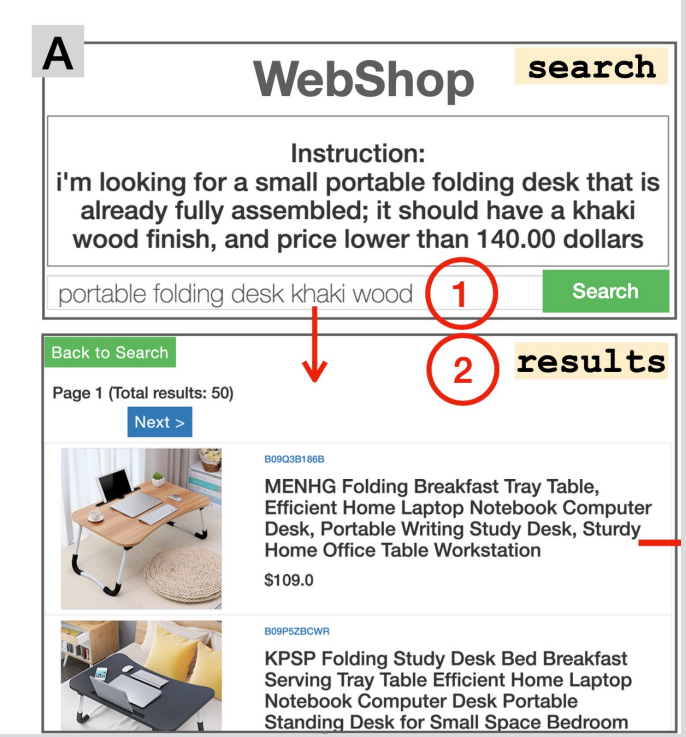


(2) Large process model as a large process model agent: example



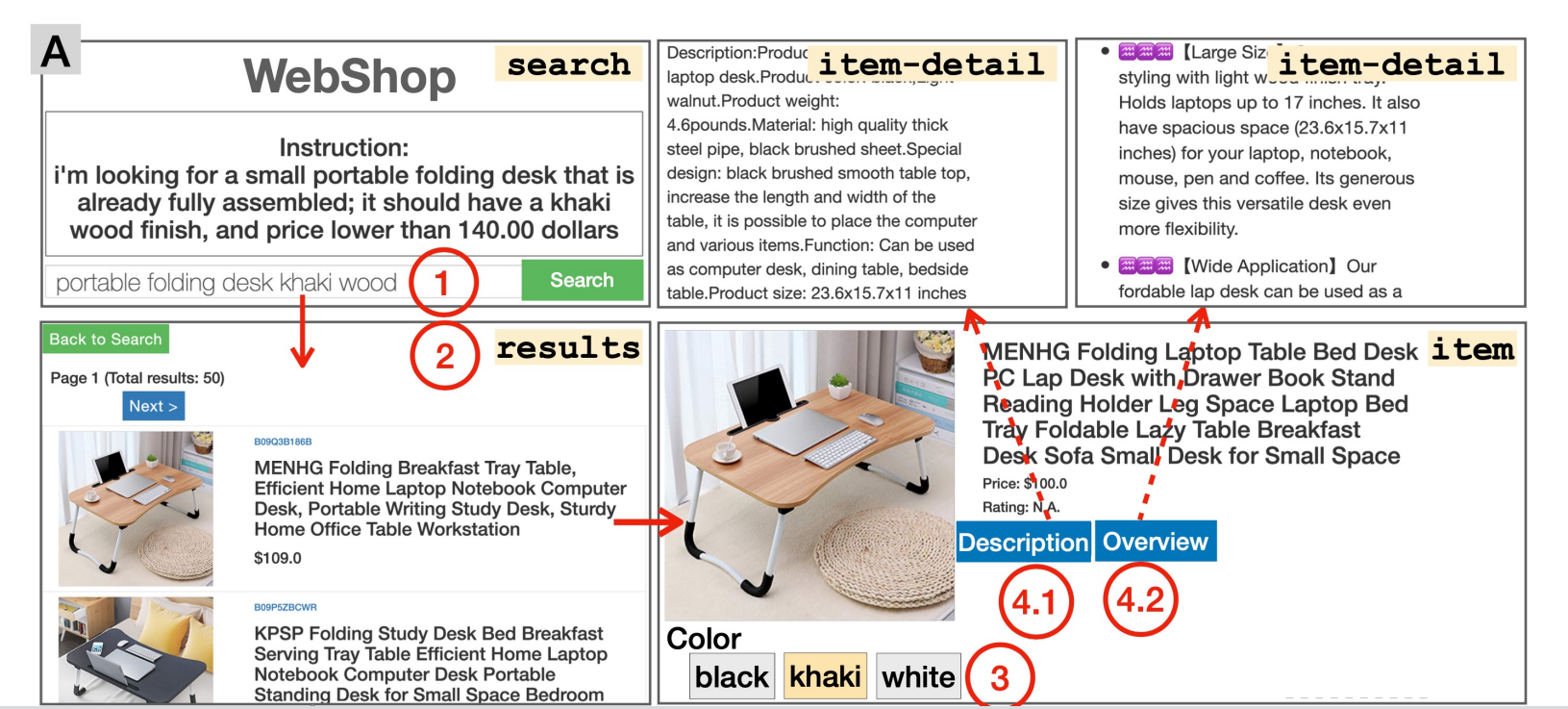
Source: Yao et al.: WebShop: Towards Scalable Real-World Web Interaction with Grounded Language Agents. arXiv:2207.01206, 2023.

(2) Large process model as a large process model agent: example



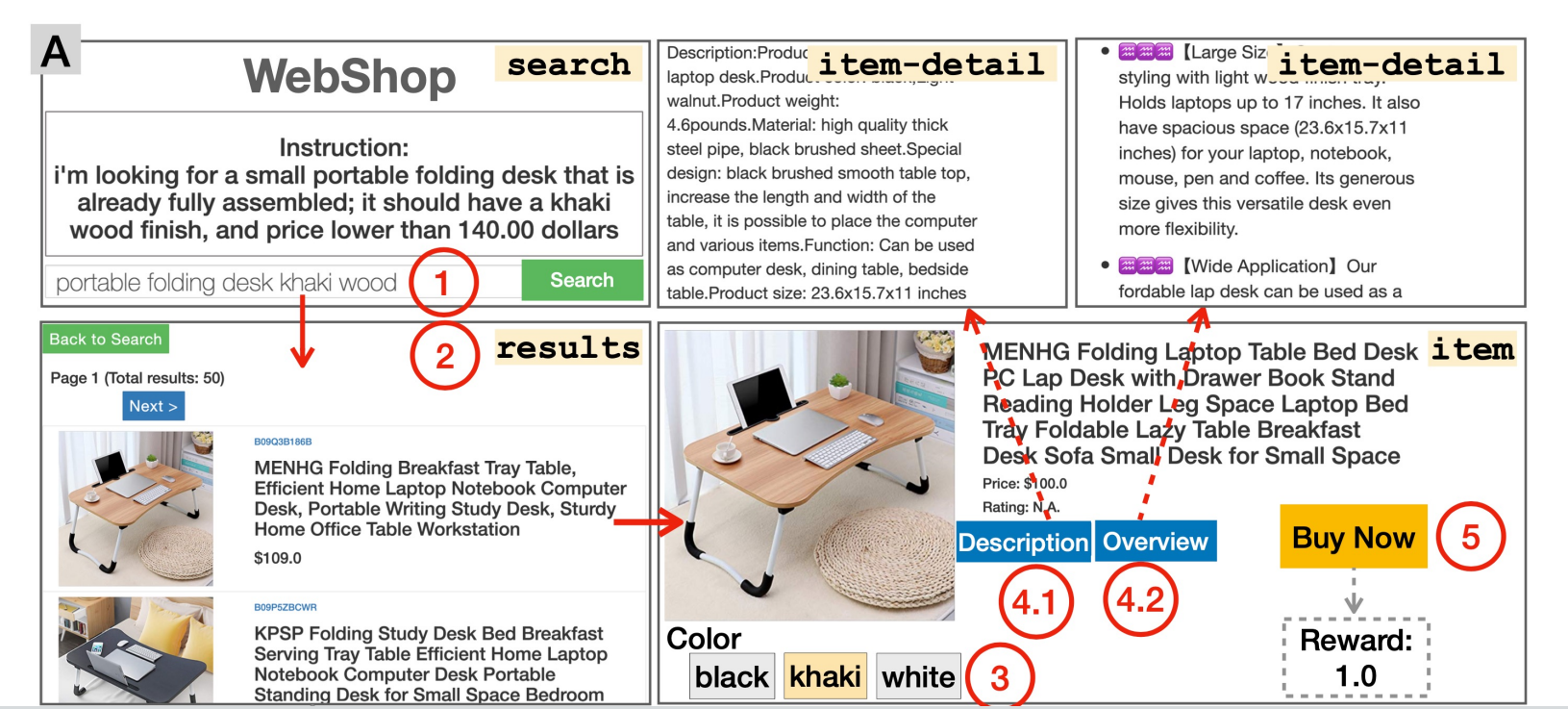
Source: Yao et al.: WebShop: Towards Scalable Real-World Web Interaction with Grounded Language Agents. arXiv:2207.01206, 2023.

(2) Large process model as a large process model agent: example



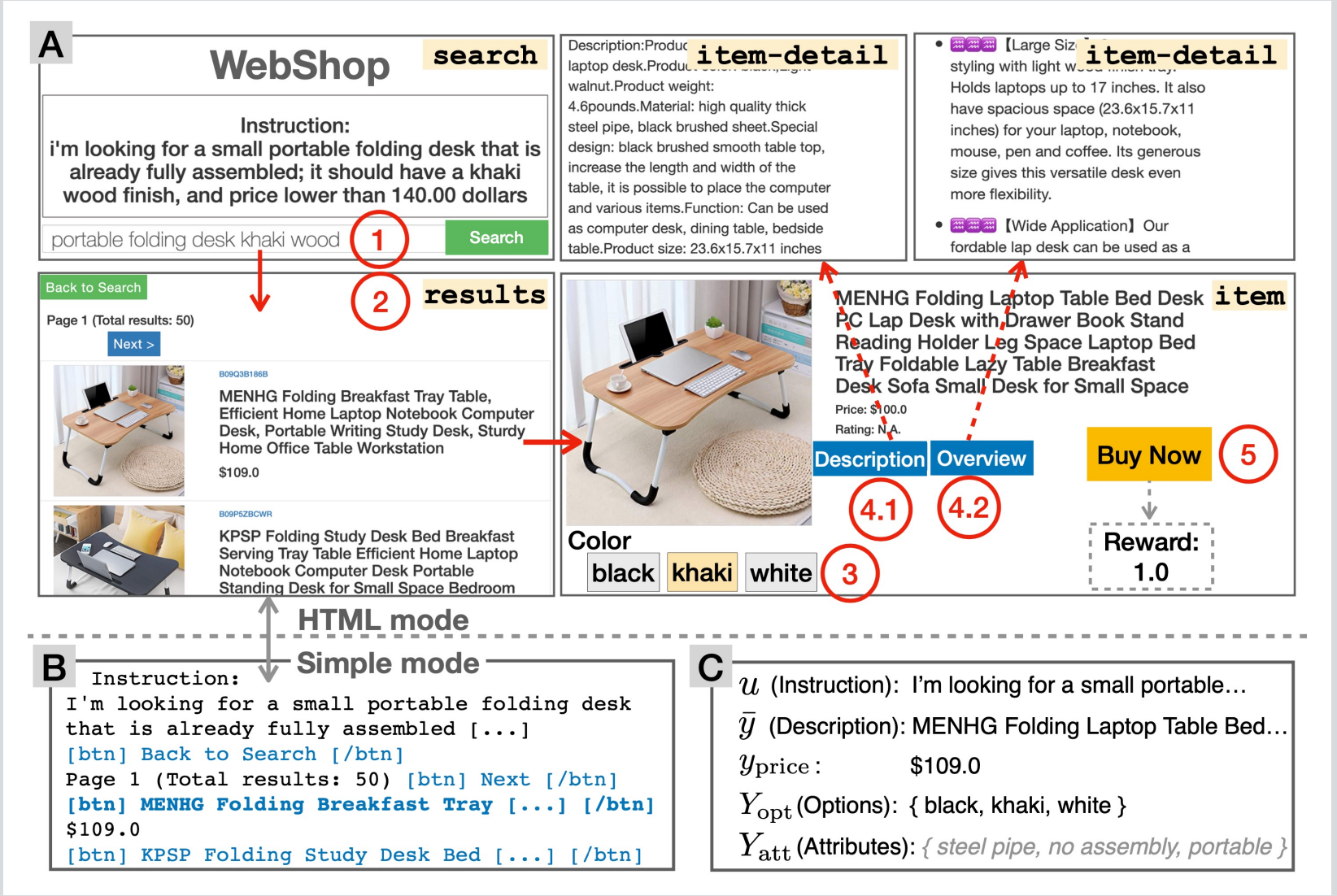
Source: Yao et al.: WebShop: Towards Scalable Real-World Web Interaction with Grounded Language Agents. arXiv:2207.01206, 2023.

(2) Large process model as a large process model agent: example



Source: Yao et al.: WebShop: Towards Scalable Real-World Web Interaction with Grounded Language Agents. arXiv:2207.01206, 2023.

(2) Large process model as a large process model agent: example



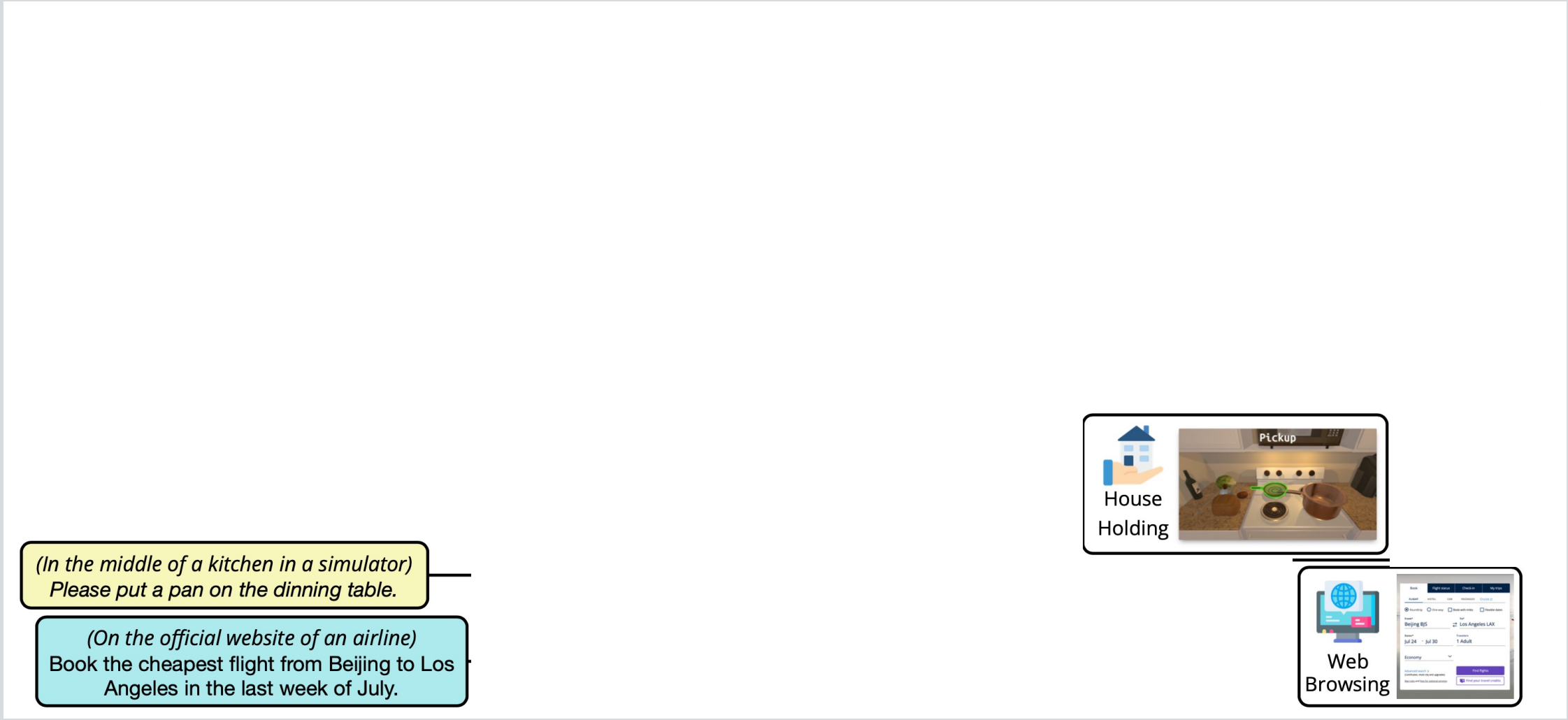
Source: Yao et al.: WebShop: Towards Scalable Real-World Web Interaction with Grounded Language Agents. arXiv:2207.01206, 2023.

(2) Large process model as a large process model agent: Evaluation benchmarks

(On the official website of an airline)
Book the cheapest flight from Beijing to Los Angeles in the last week of July.

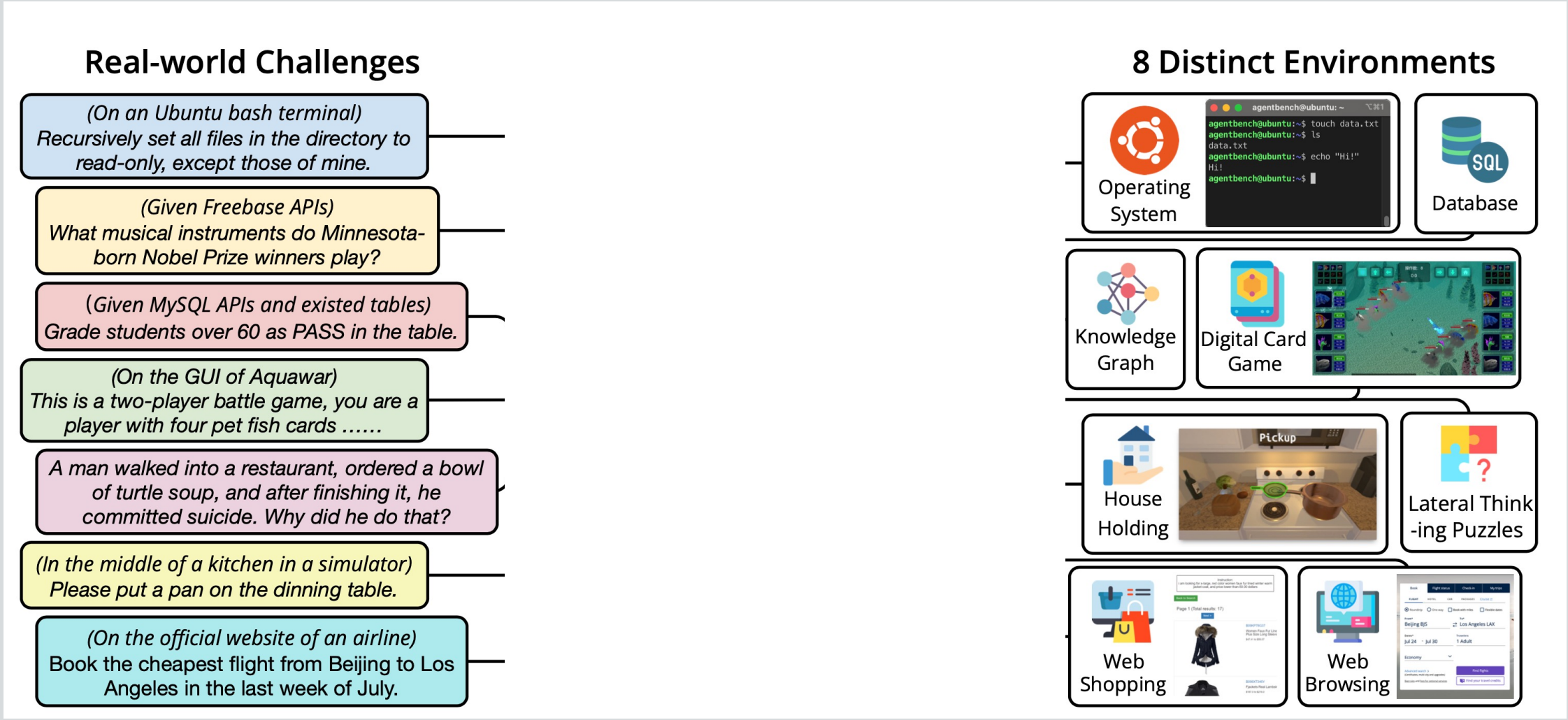
Source: Liu et al.: AgentBench: Evaluating LLMs as agents. arXiv:2308.03688, 2023,

(2) Large process model as a large process model agent: Evaluation benchmarks



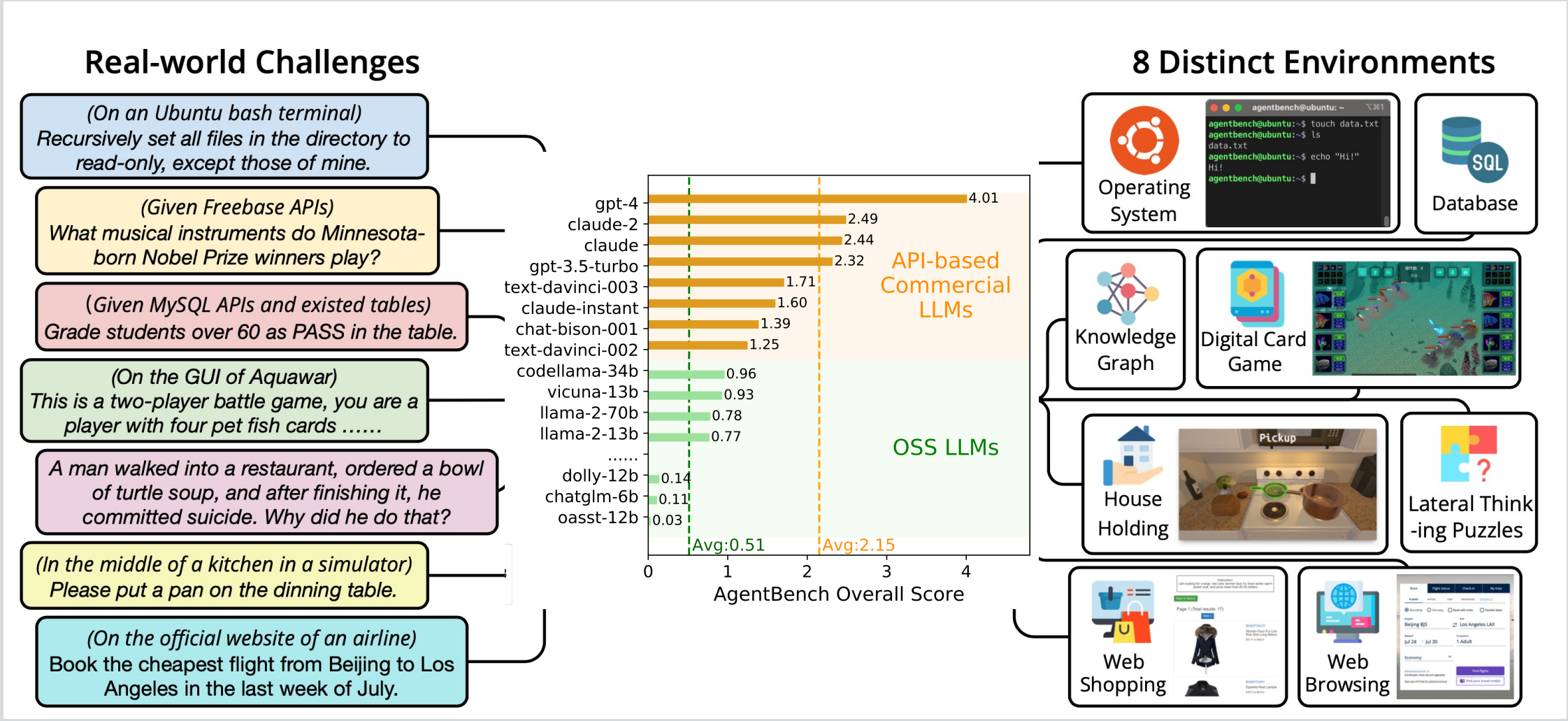
Source: Liu et al.: AgentBench: Evaluating LLMs as agents. arXiv:2308.03688, 2023,

(2) Large process model as a large process model agent: Evaluation benchmarks



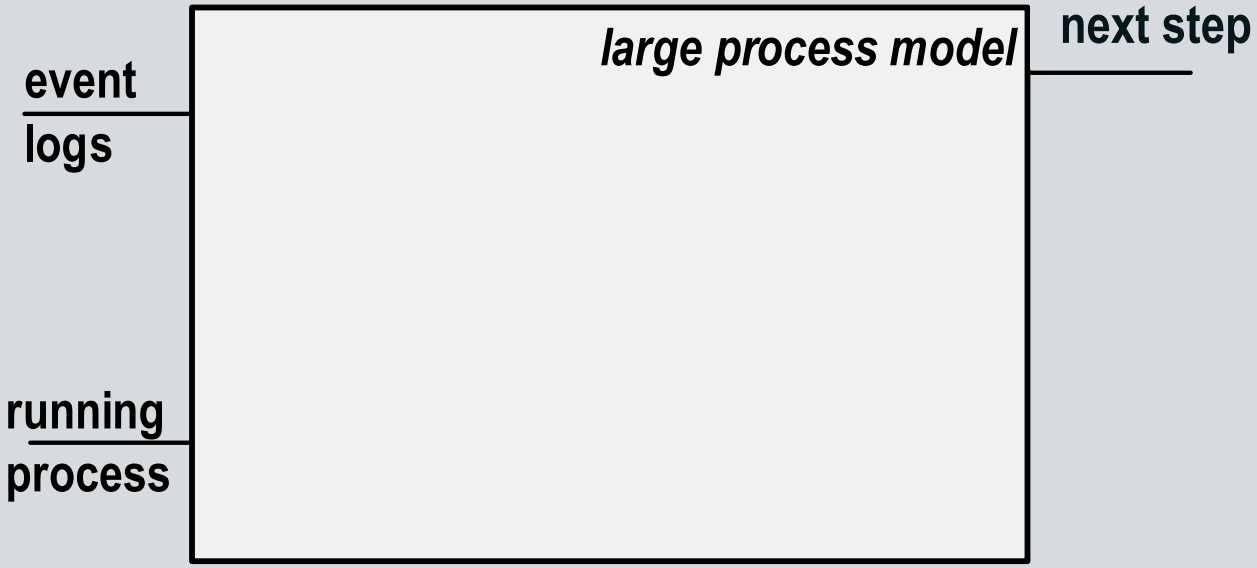
Source: Liu et al.: AgentBench: Evaluating LLMs as agents. arXiv:2308.03688, 2023,

(2) Large process model as a large process model agent: Evaluation benchmarks

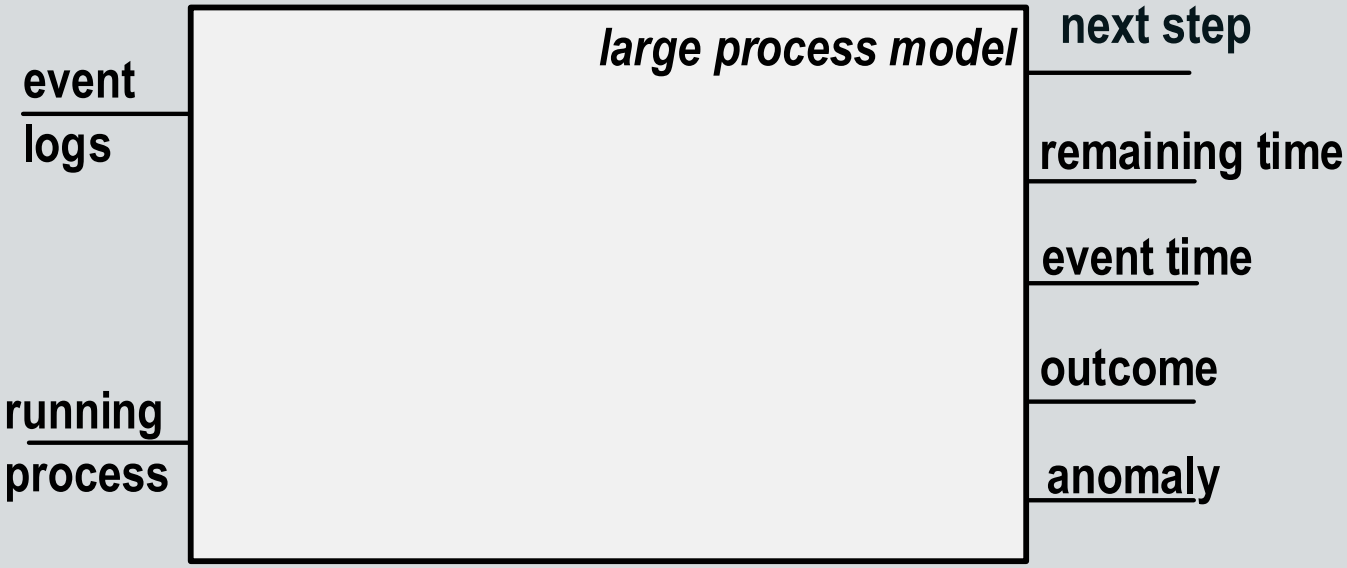


Source: Liu et al.: AgentBench: Evaluating LLMs as agents. arXiv:2308.03688, 2023,

(3) Large process model as a native large language model



(3) Large process model as a native large language model



(3) Large process model as a native large language model

Decision Support Systems 100 (2017) 129–140

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Predicting process behaviour using deep learning

Joerg Evermann[a,*], Jana-Rebecca Rehse[b,c], Peter Fettke[b,c]

[a] Memorial University of Newfoundland, St. John's, NL, Canada
[b] German Research Center for Artificial Intelligence, Saarbrücken, Germany
[c] Saarland University, Saarbrücken, Germany

ARTICLE INFO

Article history:
Received 8 July 2016
Received in revised form 22 March 2017
Accepted 5 April 2017
Available online 17 April 2017

Keywords:
Process management
Runtime support
Process prediction
Deep learning
Neural networks

ABSTRACT

Predicting business process behaviour is an important aspect of business process management. Motivated by research in natural language processing, this paper describes an application of deep learning with recurrent neural networks to the problem of predicting the next event in a business process. This is both a novel method in process prediction, which has largely relied on explicit process models, and also a novel application of deep learning methods. The approach is evaluated on two real datasets and our results surpass the state-of-the-art in prediction precision.

1. Introduction

Being able to predict the future behaviour of a business process is an important business capability [1]. As an application of predictive analytics in business process management, process prediction exploits data on past process instances to make predictions about current ones [2]. Example use cases are customer service agents responding to inquiries about the remaining time until a case is resolved, production managers predicting the completion time of a production process for better planning and higher utilization, or case managers identifying likely compliance violations to mitigate business risk.

We present a novel approach to predicting the next process event using deep learning. While the term "deep learning" has only recently become a popular research topic, it is essentially an application of neural networks and thus looks back on a long history of research [3]. Recent innovations both in algorithms, allowing novel architectures of neural networks, and computing hardware, especially GPU processing, have led to a resurgence in interest for neural networks and popularized the term "deep learning" [4]. Our approach is motivated by applications of neural networks to natural language processing(NLP), more specifically the prediction of the next word in a sentence [5–7]. By interpreting process event logs as text, process traces as sentences, and process events as words, these techniques can be applied to predict future process events. The contribution of our research is threefold:

1. We improve on the state-of-the-art in process event prediction. Our results show our method has considerably better precision on next-event prediction.
2. We demonstrate that an explicit process model is not necessary for prediction. Deep learning models, where the process structure is only implicitly reflected, can perform as well as explicit process models.
3. We contribute to process management in general by showcasing the useful application of an artificial intelligence approach, illustrating that business process management can benefit from the application of smart approaches.

Our research is located at the intersection of business process management, in particular process mining, and artificial intelligence (AI) and machine learning. We bring together historic process data with an AI learning technology to leverage real-time case management, opening new perspectives into process execution, monitoring, and analysis. Extending existing solutions to novel problems ("exaptation") is a recognized and valid way to make a contribution in design science [8], which is the research approach we apply here. We not only provide a new approach, rooted in AI, to predicting the next

* Corresponding author.
E-mail address: jevermann@mun.ca (J. Evermann).

http://dx.doi.org/10.1016/j.dss.2017.04.003
0167-9236/

(3) Large process model as a native large language model

Decision Support Systems 100 (2017) 129–140

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Predicting process behaviour using deep learning

Joerg Evermann[a,*], Jana-Rebecca Rehse[b,c], Peter Fettke[b,c]

[a] Memorial University of Newfoundland, St. John's, NL, Canada
[b] German Research Center for Artificial Intelligence, Saarbrücken, Germany
[c] Saarland University, Saarbrücken, Germany

ARTICLE INFO

Article history:
Received 8 July 2016
Received in revised form 22 March 2017
Accepted 5 April 2017
Available online 17 April 2017

Keywords:
Process management
Runtime support
Process prediction
Deep learning
Neural networks

ABSTRACT

Predicting business process behaviour is an important aspect of business process management. Motivated by research in natural language processing, this paper describes an application of deep learning with recurrent neural networks to the problem of predicting the next event in a business process. This is both a novel method in process prediction, which has largely relied on explicit process models, and also a novel application of deep learning methods. The approach is evaluated on two real datasets and our results surpass the state-of-the-art in prediction precision.

1. Introduction

Being able to predict the future behaviour of a business process is an important business capability [1]. As an application of predictive analytics in business process management, process prediction exploits data on past process instances to make predictions about current ones [2]. Example use cases are customer service agents responding to inquiries about the remaining time until a case is resolved, production managers predicting the completion time of a production process for better planning and higher utilization, or case managers identifying likely compliance violations to mitigate business risk.

We present a novel approach to predicting the next process event using deep learning. While the term "deep learning" has only recently become a popular research topic, it is essentially an application of neural networks and thus looks back on a long history of research [3]. Recent innovations both in algorithms, allowing novel architectures of neural networks, and computing hardware, especially GPU processing, have led to a resurgence in interest for neural networks and popularized the term "deep learning" [4]. Our approach is motivated by applications of neural networks to natural language processing(NLP), more specifically the prediction of the next word in a sentence [5–7]. By interpreting process event logs as text, process traces as sentences, and process events as words, these techniques can be applied to predict future process events. The contribution of our research is threefold:

1. We improve on the state-of-the-art in process event prediction. Our results show our method has considerably better precision on next-event prediction.
2. We demonstrate that an explicit process model is not necessary for prediction. Deep learning models, where the process structure is only implicitly reflected, can perform as well as explicit process models.
3. We contribute to process management in general by showcasing the useful application of an artificial intelligence approach, illustrating that business process management can benefit from the application of smart approaches.

Our research is located at the intersection of business process management, in particular process mining, and artificial intelligence (AI) and machine learning. We bring together historic process data with an AI learning technology to leverage real-time case management, opening new perspectives into process execution, monitoring, and analysis. Extending existing solutions to novel problems ("exaptation") is a recognized and valid way to make a contribution in design science [8], which is the research approach we apply here. We not only provide a new approach, rooted in AI, to predicting the next

* Corresponding author.
E-mail address: jevermann@mun.ca (J. Evermann).

http://dx.doi.org/10.1016/j.dss.2017.04.003
0167-9236/

Artificial Intelligence Review (2022) 55:801–827
https://doi.org/10.1007/s10462-021-09960-8

A systematic literature review on state-of-the-art deep learning methods for process prediction

Dominic A. Neu[1,2] · Johannes Lahann[1,2] · Peter Fettke[1,2]

Published online: 11 March 2021

Abstract
Process mining enables the reconstruction and evaluation of business processes based on digital traces in IT systems. An increasingly important technique in this context is process prediction. Given a sequence of events of an ongoing trace, process prediction allows forecasting upcoming events or performance measurements. In recent years, multiple process prediction approaches have been proposed, applying different data processing schemes and prediction algorithms. This study focuses on deep learning algorithms since they seem to outperform their machine learning alternatives consistently. Whilst having a common learning algorithm, they use different data preprocessing techniques, implement a variety of network topologies and focus on various goals such as outcome prediction, time prediction or control-flow prediction. Additionally, the set of log-data, evaluation metrics and baselines used by the authors diverge, making the results hard to compare. This paper attempts to synthesise the advantages and disadvantages of the procedural decisions in these approaches by conducting a systematic literature review.

Keywords Process prediction · Predictive process monitoring · Systematic literature review · Deep learning

1 Introduction

Today's information systems create, utilize and store vast amounts of data about the business processes being executed with them. These logs capture the as-is execution. Process mining extracts knowledge from these logs to provide means for process discovery, process monitoring and process improvement. Additionally, in case target process models are provided, conformance-checking searches for deviations of this model. Consequently, process mining is situated between the disciplines of data mining and business process modelling (Van der Aalst et al. 2011). In recent years, much effort was put into process discovery to build human-readable models for further investigation by

✉ Dominic A. Neu
dominic.neu@dfki.de

[1] Institute for Information Systems, German Research Center for Artificial Intelligence (DFKI), Saarbruecken, Germany

[2] Institute for Information Systems, Saarland University, Saarbruecken, Germany

(3) Large process model as a native large language model

Decision Support Systems 100 (2017) 129–140

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Predicting process behaviour using deep learning

Joerg Evermann[a,*], Jana-Rebecca Rehse[b,c], Peter Fettke[b,c]

[a] Memorial University of Newfoundland, St. John's, NL, Canada
[b] German Research Center for Artificial Intelligence, Saarbrücken, Germany
[c] Saarland University, Saarbrücken, Germany

ARTICLE INFO

Article history:
Received 8 July 2016
Received in revised form 22 March 2017
Accepted 5 April 2017
Available online 17 April 2017

Keywords:
Process management
Runtime support
Process prediction
Deep learning
Neural networks

ABSTRACT

Predicting business process behaviour is an important aspect of business process management. Motivated by research in natural language processing, this paper describes an application of deep learning with recurrent neural networks to the problem of predicting the next event in a business process. This is both a novel method in process prediction, which has largely relied on explicit process models, and also a novel application of deep learning methods. The approach is evaluated on two real datasets and our results surpass the state-of-the-art in prediction precision.

1. Introduction

Being able to predict the future behaviour of a business process is an important business capability [1]. As an application of predictive analytics in business process management, process prediction exploits data on past process instances to make predictions about current ones [2]. Example use cases are customer service agents responding to inquiries about the remaining time until a case is resolved, production managers predicting the completion time of a production process for better planning and higher utilization, or case managers identifying likely compliance violations to mitigate business risk.

We present a novel approach to predicting the next process event using deep learning. While the term "deep learning" has only recently become a popular research topic, it is essentially an application of neural networks and thus looks back on a long history of research [3]. Recent innovations both in algorithms, allowing novel architectures of neural networks, and computing hardware, especially GPU processing, have led to a resurgence in interest for neural networks and popularized the term "deep learning" [4]. Our approach is motivated by applications of neural networks to natural language processing(NLP), more specifically the prediction of the next word in a sentence [5–7]. By interpreting process event logs as text, process traces as sentences, and process events as words, these techniques can be applied to predict future process events. The contribution of our research is threefold:

1. We improve on the state-of-the-art in process event prediction. Our results show our method has considerably better precision on next-event prediction.
2. We demonstrate that an explicit process model is not necessary for prediction. Deep learning models, where the process structure is only implicitly reflected, can perform as well as explicit process models.
3. We contribute to process management in general by showcasing the useful application of an artificial intelligence approach, illustrating that business process management can benefit from the application of smart approaches.

Our research is located at the intersection of business process management, in particular process mining, and artificial intelligence (AI) and machine learning. We bring together historic process data with an AI learning technology to leverage real-time case management, opening new perspectives into process execution, monitoring, and analysis. Extending existing solutions to novel problems ("exaptation") is a recognized and valid way to make a contribution in design science [8], which is the research approach we apply here. We not only provide a new approach, rooted in AI, to predicting the next

* Corresponding author.
E-mail address: jevermann@mun.ca (J. Evermann).

http://dx.doi.org/10.1016/j.dss.2017.04.003
0167-9236/

Artificial Intelligence Review (2022) 55:801–827
https://doi.org/10.1007/s10462-021-09960-8

A systematic literature review on state-of-the-art deep learning methods for process prediction

Dominic A. Neu[1,2] · Johannes Lahann[1,2] · Peter Fettke[1,2]

Published online: 11 March 2021

Abstract
Process mining enables the reconstruction and evaluation of business processes based on digital traces in IT systems. An increasingly important technique in this context is process prediction. Given a sequence of events of an ongoing trace, process prediction allows forecasting upcoming events or performance measurements. In recent years, multiple process prediction approaches have been proposed, applying different data processing schemes and prediction algorithms. This study focuses on deep learning algorithms since they seem to outperform their machine learning alternatives consistently. Whilst having a common learning algorithm, they use different data preprocessing techniques, implement a variety of network topologies and focus on various goals such as outcome prediction, time prediction or control-flow prediction. Additionally, the set of log-data, evaluation metrics and baselines used by the authors diverge, making the results hard to compare. This paper attempts to synthesise the advantages and disadvantages of the procedural decisions in these approaches by conducting a systematic literature review.

Keywords Process prediction · Predictive process monitoring · Systematic literature review · Deep learning

1 Introduction

Today's information systems create, utilize and store vast amounts of data about the business processes being executed with them. These logs capture the as-is execution. Process mining extracts knowledge from these logs to provide means for process discovery, process monitoring and process improvement. Additionally, in case target process models are provided, conformance-checking searches for deviations of this model. Consequently, process mining is situated between the disciplines of data mining and business process modelling (Van der Aalst et al. 2011). In recent years, much effort was put into process discovery to build human-readable models for further investigation by

✉ Dominic A. Neu
dominic.neu@dfki.de

[1] Institute for Information Systems, German Research Center for Artificial Intelligence (DFKI), Saarbruecken, Germany

[2] Institute for Information Systems, Saarland University, Saarbruecken, Germany

PGTNet: A Process Graph Transformer Network for Remaining Time Prediction of Business Process Instances

Keyvan Amiri Elyasi[1[0009−0007−3016−2392]], Han van der Aa[2[0000−0002−4200−4937]], and Heiner Stuckenschmidt[1[0000−0002−0209−3859]]

[1] Data and Web Science Group, University of Mannheim, Germany
{keyvan,heiner}@informatik.uni-mannheim.de
[2] Faculty of Computer Science, University of Vienna, Austria
han.van.der.aa@univie.ac.at

Abstract. We present PGTNet, an approach that transforms event logs into graph datasets and leverages graph-oriented data for training Process Graph Transformer Networks to predict the remaining time of business process instances. PGTNet consistently outperforms state-of-the-art deep learning approaches across a diverse range of 20 publicly available real-world event logs. Notably, our approach is most promising for highly complex processes, where existing deep learning approaches encounter difficulties stemming from their limited ability to learn control-flow relationships among process activities and capture long-range dependencies. PGTNet addresses these challenges, while also being able to consider multiple process perspectives during the learning process.

Keywords: Predictive process monitoring · Remaining time prediction · Deep learning · Graph Transformers.

1 Introduction

Predictive process monitoring (PPM) aims to forecast the future behaviour of running business process instances, thereby enabling organizations to optimize their resource allocation and planning [17], as well as take corrective actions [7]. An important task in PPM is remaining time prediction, which strives to accurately predict the time until an active process instance will be completed. Precise estimations for remaining time are crucial for avoiding deadline violations, optimizing operational efficiency, and providing estimates to customers [13, 17].

A variety of approaches have been developed to tackle remaining time prediction, with recent works primarily being based on deep learning architectures. In this regard, approaches using deep neural networks are among the most prominent ones [15]. However, the predictive accuracy of these networks leaves considerable room for improvement. In particular, they face challenges when it comes to capturing long-range dependencies [2] and other control-flow relationships (such as loops and parallelism) between process activities [22], whereas they

(3) Large … model

Table 2. Mean Absolute Error for remaining time prediction (MAE: in days).

Event log	DUMMY	DALSTM	Process Transformer	GGNN	PGTNet
Env.permit	5.21	3.36± 0.04	4.26± 0.04	3.52± 0.02	**2.72± 0.08**
Helpdesk	9.15	8.22± 0.23	6.33± 0.01	6.21± 0.04	**4.11± 0.04**
BPIC12	9.03	9.34± 0.41	7.11± 0.02	4.78± 0.01	**2.31± 0.19**
BPIC12W	9.16	8.22± 0.06	7.40± 0.01	5.12± 0.02	**2.70± 0.01**
BPIC12C	8.92	8.21± 0.27	6.86± 0.01	5.32± 0.01	**2.77± 0.02**
BPIC12CW	9.17	8.04± 0.09	7.46± 0.01	6.99± 0.01	**5.07± 0.03**
BPIC12O	8.39	8.21± 0.09	7.29± 0.01	6.93± 0.04	**5.57± 0.01**
BPIC12A	8.17	7.62± 0.03	7.79± 0.01	7.48± 0.01	**7.38± 0.01**
BPIC20I	27.20	20.43± 0.39	17.06± 0.11	15.67± 0.04	**7.67± 0.19**
BPIC20D	4.33	4.15± 0.12	3.65± 0.01	3.25± 0.01	**3.10± 0.01**
Sepsis	41.12	25.21± 0.66	34.77± 0.18	19.44± 0.05	**16.48± 0.19**
Hospital	59.41	43.66± 0.10	47.00± 0.07	41.84± 0.06	**35.68± 0.03**
BPIC15-1	50.22	36.48± 2.69	31.01± 0.36	16.77± 0.01	**1.76± 0.06**
BPIC15-2	83.11	63.66± 2.36	44.04± 0.48	20.76± 0.05	**3.02± 0.07**
BPIC15-3	28.76	17.69± 1.16	15.23± 0.23	7.06± 0.03	**1.54± 0.23**
BPIC15-4	56.75	53.33± 2.63	34.40± 0.42	17.97± 0.03	**1.65± 0.06**
BPIC15-5	45.97	42.89± 3.08	27.76± 0.28	13.61± 0.08	**1.61± 0.01**
BPIC13I	16.18	7.60± 0.45	13.54± 0.04	11.99± 0.04	**2.23± 0.05**
BPIC13C	152.93	91.82± 1.48	127.01± 0.85	123.28± 0.53	**37.44± 1.49**
Traffic fines	196.26	187.41± 0.53	187.08± 0.11	154.56± 0.19	**113.53± 0.12**
Average	41.47	32.78± 0.84	31.85± 0.16	24.63± 0.06	**12.92± 0.14**

Decision Support Systems 1…
Contents lists available …
Decision Suppo…
journal homepage: www.el…

Predicting process behaviour using deep lea…

Joerg Evermann[a,*], Jana-Rebecca Rehse[b,c], Peter Fettke[b]
[a] Memorial University of Newfoundland, St. John's, NL, Canada
[b] German Research Center for Artificial Intelligence, Saarbrücken, Germany
[c] Saarland University, Saarbrücken, Germany

ARTICLE INFO
Article history:
Received 8 July 2016
Received in revised form 22 March 2017
Accepted 5 April 2017
Available online 17 April 2017
Keywords:
Process management
Runtime support
Process prediction
Deep learning
Neural networks

ABSTRACT
Predicting business process beha…
by research in natural language p…
rent neural networks to the prob…
method in process prediction, wh…
tion of deep learning methods. T…
state-of-the-art in prediction pre…

1. Introduction

Being able to predict the future behaviour of a business process is an important business capability [1]. As an application of predictive analytics in business process management, process prediction exploits data on past process instances to make predictions about current ones [2]. Example use cases are customer service agents responding to inquiries about the remaining time until a case is resolved, production managers predicting the completion time of a production process for better planning and higher utilization, or case managers identifying likely compliance violations to mitigate business risk.

We present a novel approach to predicting the next process event using deep learning. While the term "deep learning" has only recently become a popular research topic, it is essentially an application of neural networks and thus looks back on a long history of research [3]. Recent innovations both in algorithms, allowing novel architectures of neural networks, and computing hardware, especially GPU processing, have led to a resurgence in interest for neural networks and popularized the term "deep learning" [4]. Our approach is motivated by applications of neural networks to natural language processing(NLP), more specifically the prediction of the next word in a sentence [5–7]. By interpreting process event logs …

* Corresponding author.
E-mail address: jevermann@mun.ca (J. Evermann).

http://dx.doi.org/10.1016/j.dss.2017.04.003
0167-9236/

… Transformer Network
…diction of Business
…tances

…-3016−2392], Han van der
…kenschmidt[1][0000−0002−0209−3859]
…rsity of Mannheim, Germany
…k.uni-mannheim.de
…iversity of Vienna, Austria
…ivie.ac.at

…roach that transforms event logs
…-oriented data for training Pro-
…dict the remaining time of busi-
…tly outperforms state-of-the-art
…e range of 20 publicly available
…ach is most promising for highly
… learning approaches encounter
…ibility to learn control-flow rela-
…pture long-range dependencies.
…ile also being able to consider
… learning process.

…ng · Remaining time prediction

… to forecast the future behaviour of
… enabling organizations to optimize
…s well as take corrective actions [7].
…e prediction, which strives to accu-
…s instance will be completed. Precise
…r avoiding deadline violations, opti-
… estimates to customers [13, 17].
…ped to tackle remaining time predic-
…d on deep learning architectures. In
…etworks are among the most promi-
…racy of these networks leaves consid-
… they face challenges when it comes
…nd other control-flow relationships
…rocess activities [22], whereas they

Central theorem (based on [Fettke / Reisig 2024])

Each finite process R can be composed as $R = P_1 \bullet \ldots \bullet P_n$ from steps $P_1, \ldots, P_n$.

Steps $P_1, \ldots, P_n$ can be interpreted as *tokens* (“the vocabulary of a system”) and conventional methods for large language models can be employed.

Conclusions

(A) large language models evolve into a foundational technology

Conclusions

(A) large language models evolve into a foundational technology

(B) three approaches for large process models:
- (1) fine-tuned large language models
- (2) large language model agents
- (3) native large process models

Conclusions

(A) large language models evolve into a foundational technology

(B) three approaches for large process models:
    (1) fine-tuned large language models
    (2) large language model agents
    (3) native large process models

(C) composition calculus is the theoretical foundation for large process models

For a small obolus to Wolfgang, I will gladly answer your question!

Professor Dr. Peter Fettke

Saarland University and
German Research Center for Artificial Intelligence (DFKI)
Campus D 3 2
66123 Saarbrücken, Germany
peter.fettke@dfki.de, http://bpm.dfki.de
Phone: +49 681 85775-5142

dfki ai Deutsches Forschungszentrum für Künstliche Intelligenz
German Research Center for Artificial Intelligence

UNIVERSITÄT DES SAARLANDES