GIVUP: Automated Generation and Verification of Textual Process Descriptions

NIVON Quentin, SALAÜN Gwen, LANG Frédéric



Introduction

What is **BPMN**?



A workflow-based notation created in 2004 by the Business Process Management Initiative (BPMI) and the Object Management Group (OMG).

- It aims at representing business processes in a way that is understandable for both experienced and novice users.
- > An **ISO/IEC standard** since version 2.0 in 2013.



Introduction

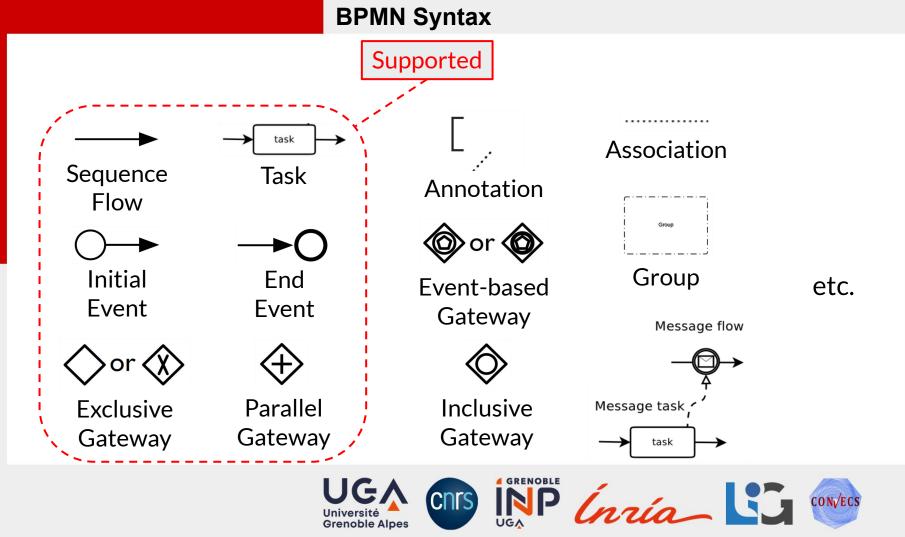
What is **BPMN**?



A workflow-based notation created in 2004 by the Business Process Management Initiative (BPMI) and the Object Management Group (OMG).

- It aims at representing business processes in a way that is understandable for both experienced and novice users.
- > An **ISO/IEC standard** since version 2.0 in 2013.





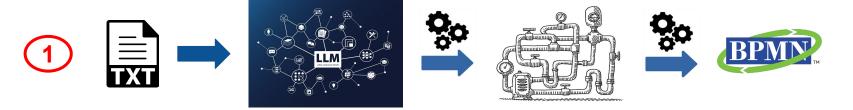
Context

- Companies are making use of the BPMN notation to represent their business processes.
- They hire experts to analyse and design the most adequate BPMN process according to their needs.
 They also expect the experts to build a process having the desired behaviour.
- These processes are often syntactically/semantically incorrect.
 Often, they do not meet the expected requirements.

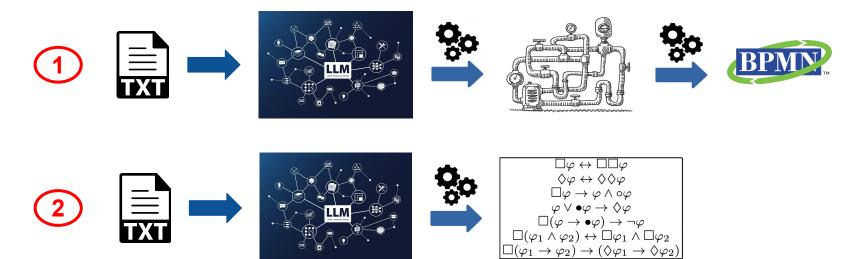


- > What if you do not know **how to write BPMN**?
- What if you do not want to spend time designing your BPMN process graphically?
- How can you be sure that your BPMN process is syntactically/semantically correct?
- How can you be sure that your BPMN process has the expected behaviour?

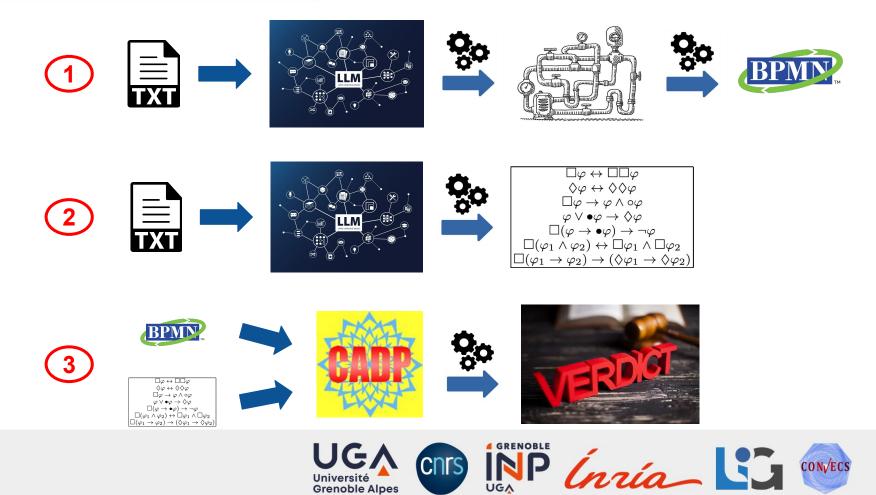






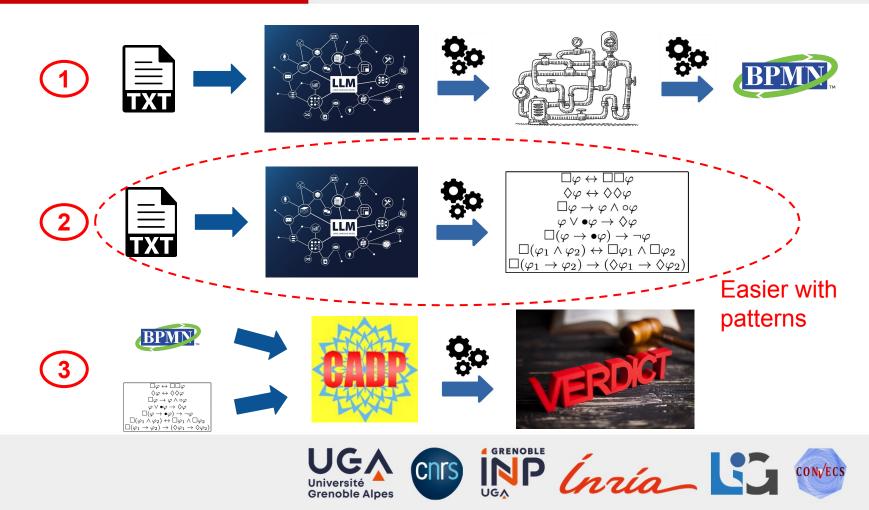






CNrs

Université **Grenoble Alpes** CON/ECS



First of all, an employee CollectGoods. Then, the client PayForDelivery while the employee PrepareParcel. Finally, the company can either DeliverByCar or DeliverByDrone (depending on the distance for example)

Textual Representation of the Process



First of all, an employee CollectGoods. Then, the client PayForDelivery while the employee PrepareParcel. Finally, the company can either DeliverByCar or DeliverByDrone (depending on the distance for example)

Textual Representation of the Process



Large Language Model (LLM)



First of all, an employee CollectGoods. Then, the client PayForDelivery while the employee PrepareParcel. Finally, the company can either DeliverByCar or DeliverByDrone (depending on the distance for example)

Textual Representation of the Process



Large Language Model (LLM) CollectGoods < (PayForDelivery, PrepareParcel)
 (PayForDelivery, PrepareParcel) < (DeliverByCar, DeliverByDrone)

$\langle E \rangle ::=$	t \mid $(\langle E \rangle) \mid$
	$\langle E_1 \rangle \langle op \rangle \langle E_2 \rangle \mid (\langle E_1 \rangle) *$
$\langle \mathrm{op} \rangle ::=$	(, ,,,,,,,

Expressions Following an Internal Grammar



First of all, an employee CollectGoods. Then, the client PayForDelivery while the employee PrepareParcel. Finally, the company can either DeliverByCar or DeliverByDrone (depending on the distance for example)

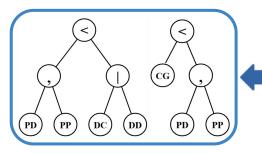
Textual Representation of the Process



Large Language Model (LLM) CollectGoods < (PayForDelivery, PrepareParcel)
 (PayForDelivery, PrepareParcel) < (DeliverByCar, DeliverByDrone)

 $\langle E \rangle ::=$ $(\langle E \rangle)$ t $\langle E_1 \rangle \langle op \rangle \langle E_2 \rangle$ $| (\langle E_1 \rangle) *$ <'ډ , '&' $\langle op \rangle ::=$

Expressions Following an Internal Grammar

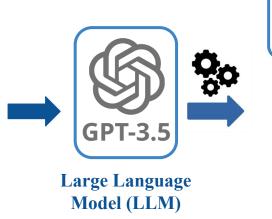


Abstract Syntax Trees



First of all, an employee CollectGoods. Then, the client PayForDelivery while the employee PrepareParcel. Finally, the company can either DeliverByCar or DeliverByDrone (depending on the distance for example)

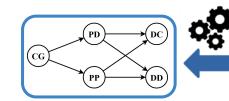
Textual Representation of the Process



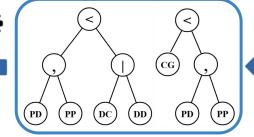
CollectGoods < (PayForDelivery, PrepareParcel)
 (PayForDelivery, PrepareParcel) < (DeliverByCar, DeliverByDrone)

 $\langle E \rangle ::=$ $(\langle E \rangle)$ $| (\langle E_1 \rangle) *$ $\langle E_1 \rangle \langle op \rangle \langle E_2 \rangle$ '&' <'ډ , $\langle op \rangle ::=$

Expressions Following an Internal Grammar



Dependency Graph



Abstract Syntax Trees





CON/ECS

First of all, an employee CollectGoods. Then, the client PayForDelivery while the employee PrepareParcel. Finally, the company can either DeliverByCar or DeliverByDrone (depending on the distance for example)

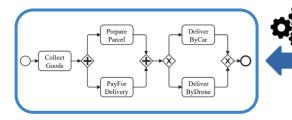
Textual Representation of the Process



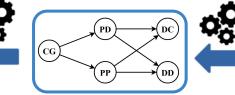
Large Language Model (LLM) CollectGoods < (PayForDelivery, PrepareParcel)
 (PayForDelivery, PrepareParcel) < (DeliverByCar, DeliverByDrone)

 $\langle E \rangle ::=$ $(\langle E \rangle)$ $| (\langle E_1 \rangle) *$ $\langle E_1 \rangle \langle op \rangle \langle E_2 \rangle$ '&' ' < 'ډ , $\langle op \rangle ::=$

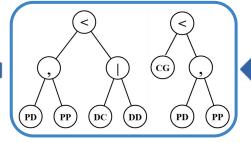
Expressions Following an Internal Grammar



BPMN Process



Dependency Graph



Abstract Syntax Trees









The user first has to write a **textual description** of the process-to-be.

First, the banker either <u>CreateProfile</u> (CP) for the user, or, if it is not needed, he <u>RetrieveCustomerProfile</u> (RCP) which triggers the system to perform the <u>AnalyseCustomerProfile</u> (ACP) task. Then, the user executes the task <u>ReceiveSupportDocuments</u> (RSD) so that the system can start <u>UpdateInfoRecords</u> (UID) and perform a <u>BackgroundVerification</u> (BV). If the verification finds missing or incorrect information, the system <u>RequestAdditionalInfo</u> (RAI) to the user, who has to <u>ReceiveSupportDocuments</u> (RSD) again. Otherwise, the process ends with <u>CreateAccount</u> (CA).



The textual description is then given to a (fine-tuned) LLM (GPT 3.5 atm).

First, the banker either $\underline{\mbox{CreateProfile}}$ (CP) for the user, or, if it is not

needed, he <u>RetrieveCustomerProfile</u> (RCP) which triggers the system to perform the <u>AnalyseCustomerProfile</u> (ACP) task. Then, the user executes the task <u>ReceiveSupportDocuments</u> (RSD) so that the system can start <u>UpdateInfoRecords</u> (UID) and perform a <u>BackgroundVerification</u> (BV).

If the verification finds missing or incorrect information, the system <u>RequestAdditionalInfo</u> (RAI) to the user, who has to <u>ReceiveSupportDocuments</u> (RSD) again.

Otherwise, the process ends with CreateAccount (CA).





GPT - 3.5



The textual description is then given to a (fine-tuned) LLM (GPT 3.5 atm).

First, the banker either $\underline{\mbox{CreateProfile}}$ (CP) for the user, or, if it is not

needed, he <u>RetrieveCustomerProfile</u> (RCP) which triggers the system to perform the <u>AnalyseCustomerProfile</u> (ACP) task. Then, the user executes the task <u>ReceiveSupportDocuments</u> (RSD) so that the system can start <u>UpdateInfoRecords</u> (UID) and perform a <u>BackgroundVerification</u> (BV).

If the verification finds missing or incorrect information, the system <u>RequestAdditionalInfo</u> (RAI) to the user, who has to <u>ReceiveSupportDocuments</u> (RSD) again.

Otherwise, the process ends with CreateAccount (CA).





The LLM processes the description and returns a **set of expressions** following an **internal grammar**.



Given our description, the LLM returns three expressions:



Given our description, the LLM returns three expressions:

(RetrieveCustomerProfile < AnalyseCustomerProfile) | CreateProfile



Given our description, the LLM returns three expressions:

(RetrieveCustomerProfile < AnalyseCustomerProfile) | CreateProfile

(RetrieveCustomerProfile, AnalyseCustomerProfile, CreateProfile) <
(ReceiveSupportDocuments < (UpdateInfoRecords, BackgroundVerification))</pre>



Given our description, the LLM returns three expressions:

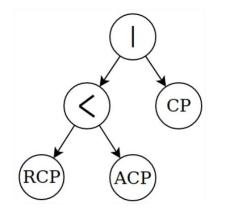
(RetrieveCustomerProfile < AnalyseCustomerProfile) | CreateProfile

(RetrieveCustomerProfile, AnalyseCustomerProfile, CreateProfile) <
(ReceiveSupportDocuments < (UpdateInfoRecords, BackgroundVerification))</pre>

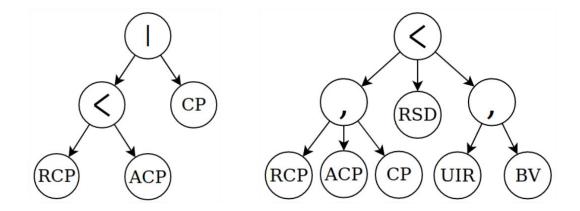
(UpdateInfoRecords, BackgroundVerification) < ((RequestAdditionalInfo < ReceiveSupportDocuments) | CreateAccount)



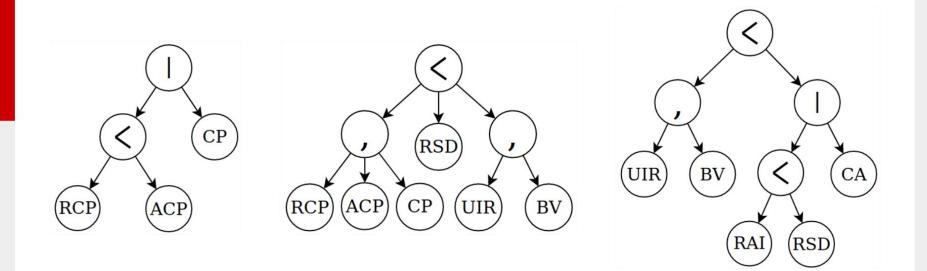






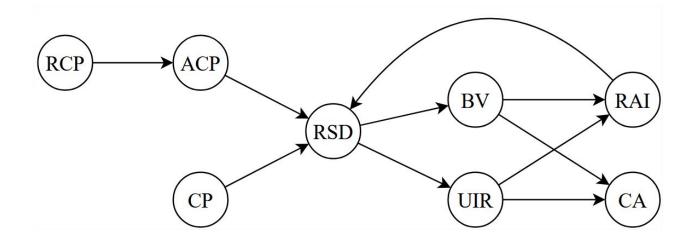








The **sequential information** contained in the multiple ASTs is gathered to obtain a **cleaner** representation of it, called **dependency graph**.

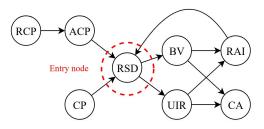




If the dependency graph **contains loops**, they are **analysed**, and all the **information needed to reconstruct** them is extracted.



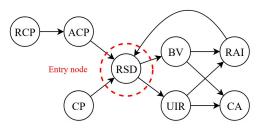
If the dependency graph **contains loops**, they are **analysed**, and all the **information needed to reconstruct** them is extracted.



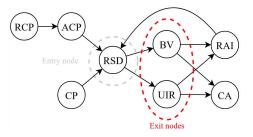
Entry node(s) computation



If the dependency graph **contains loops**, they are **analysed**, and all the **information needed to reconstruct** them is extracted.



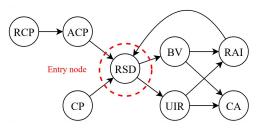
Entry node(s) computation



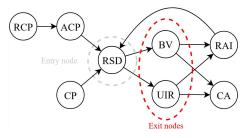
Exit node(s) computation



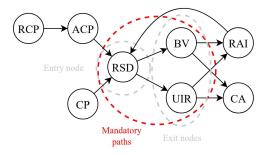
If the dependency graph **contains loops**, they are **analysed**, and all the **information needed to reconstruct** them is extracted.



Entry node(s) computation



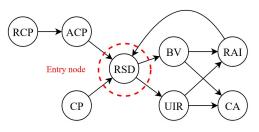
Exit node(s) computation



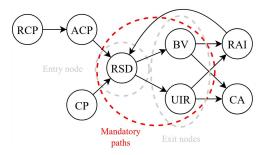
Mandatory path(s) computation



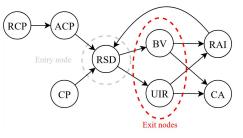
If the dependency graph **contains loops**, they are **analysed**, and all the **information needed to reconstruct** them is extracted.



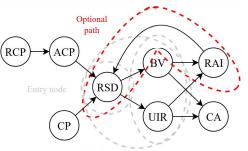
Entry node(s) computation



Mandatory path(s) computation



Exit node(s) computation

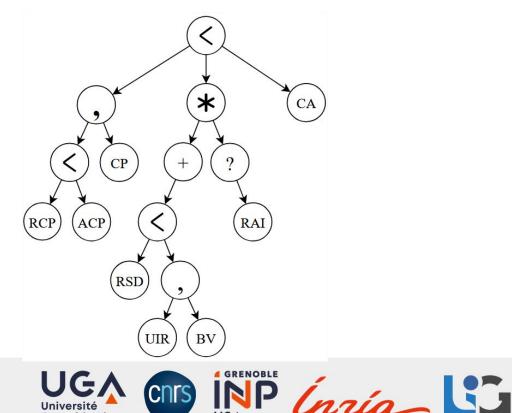


Optional path(s) computation



Once the loops have been retrieved, the **AST corresponding to the dependency graph** is built from the dependency graph.

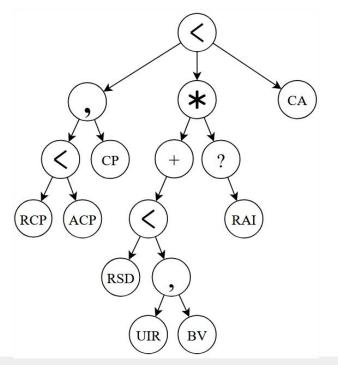
Grenoble Alpes



ŪGA

CON/ECS

The **ultimate step** to obtain an AST containing all the information belonging to the original expressions consists in **inserting the choices**.

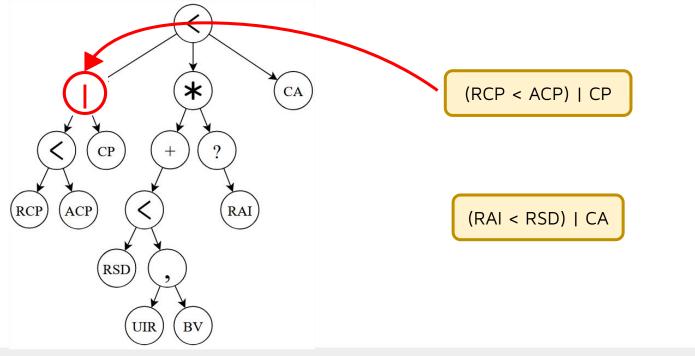




(RAI < RSD) | CA



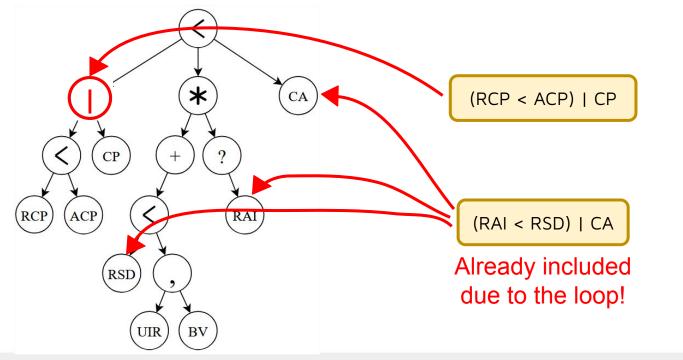
The **ultimate step** to obtain an AST containing all the information belonging to the original expressions consists in **inserting the choices**.





Detailed Solution of (1) – Step 8

The **ultimate step** to obtain an AST containing all the information belonging to the original expressions consists in **inserting the choices**.

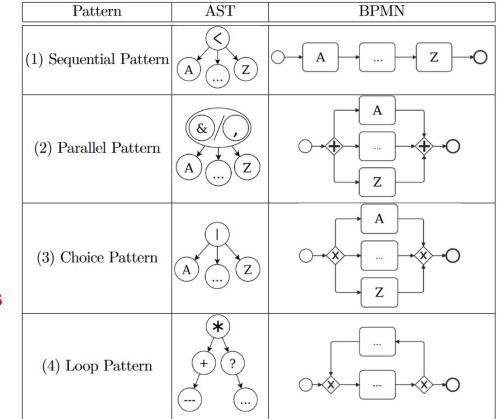




Detailed Solution of (1) – Step 9

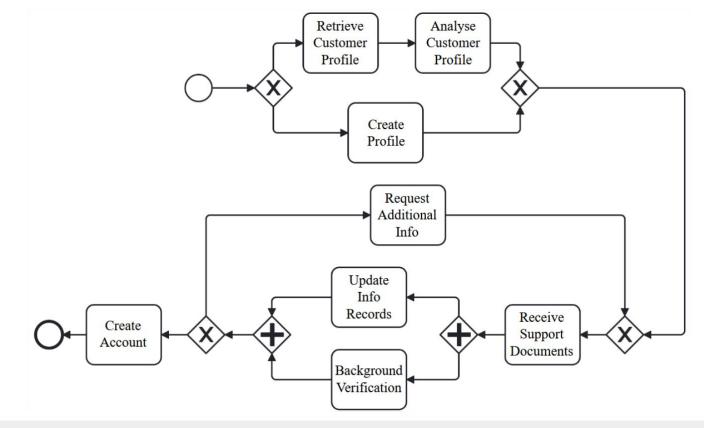
Now that the AST contains the choices, the **BPMN process is** ready to be generated.

To do so, **patterns** are applied recursively to the merged AST, **starting from the deepest nodes** (leafs).





Detailed Solution of (1) – Result





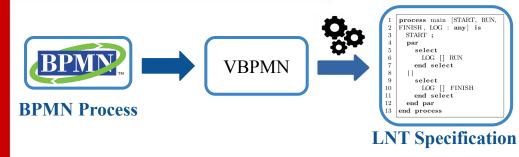


BPMN Process

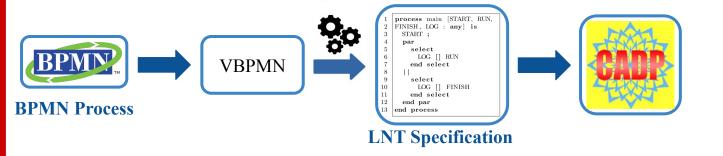




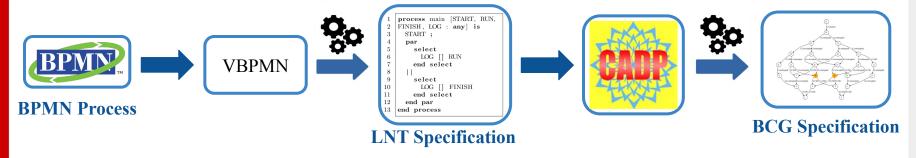




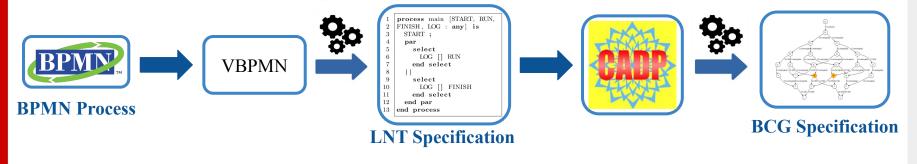








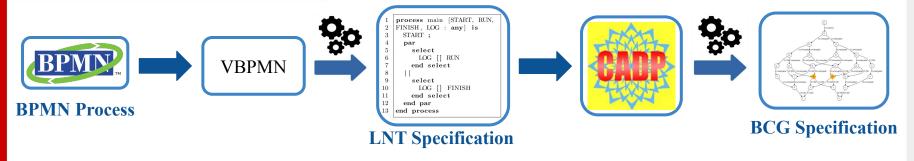






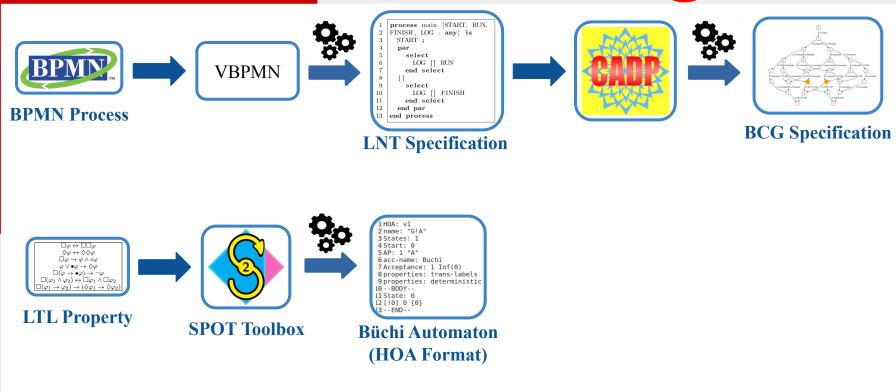
LTL Property



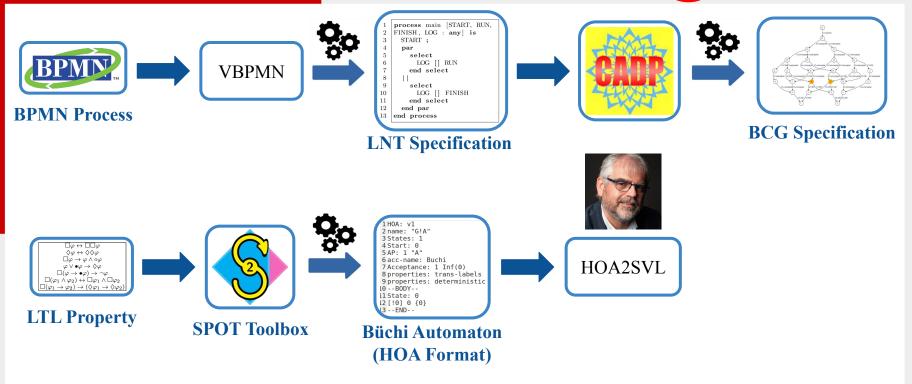




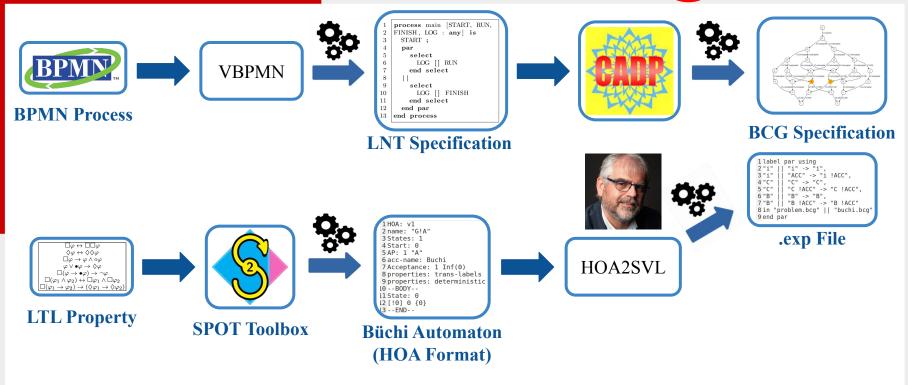




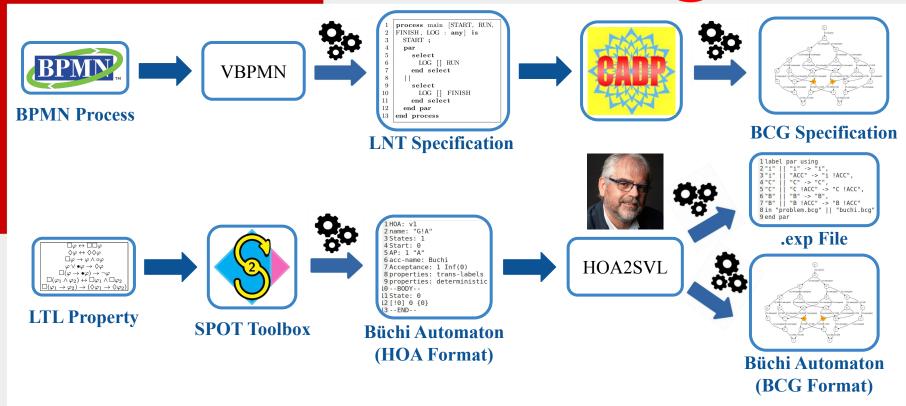




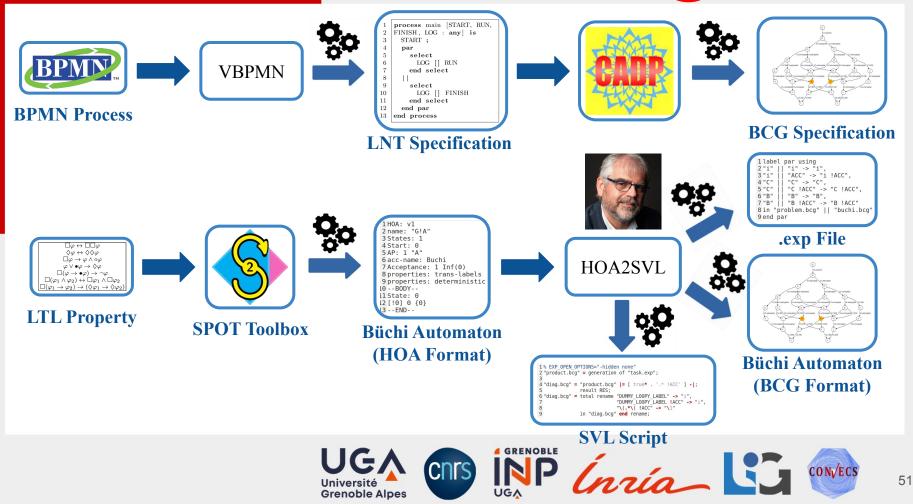








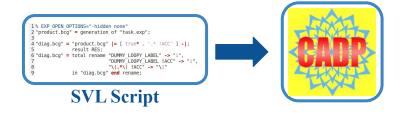




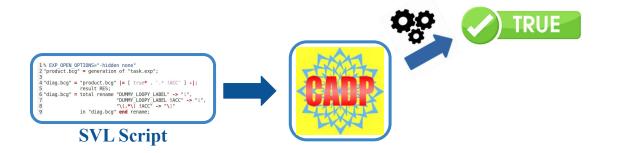




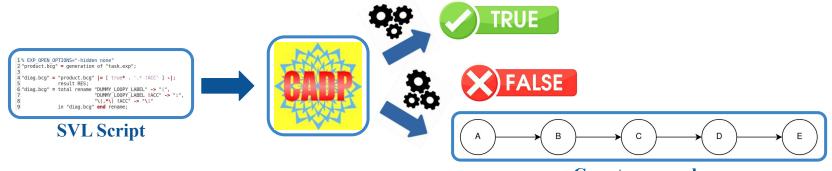






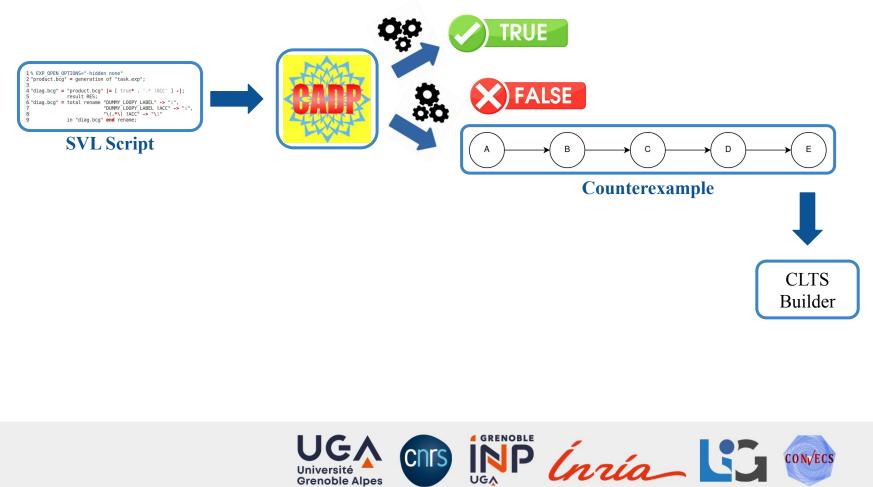


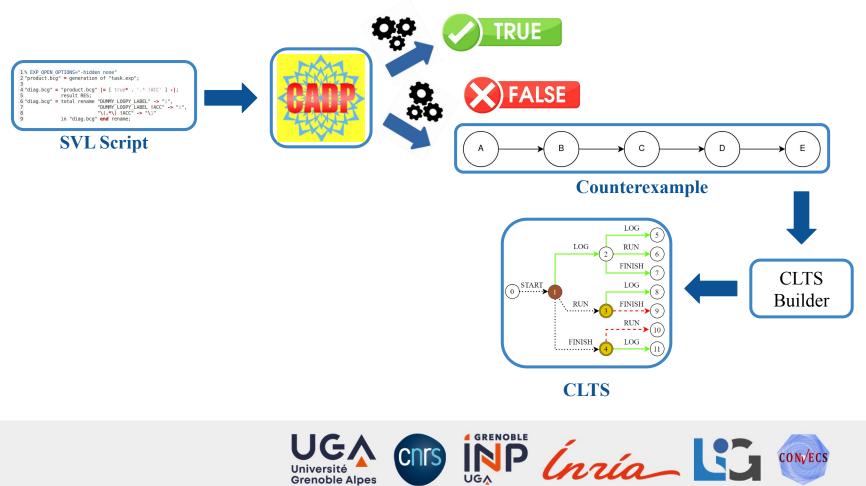




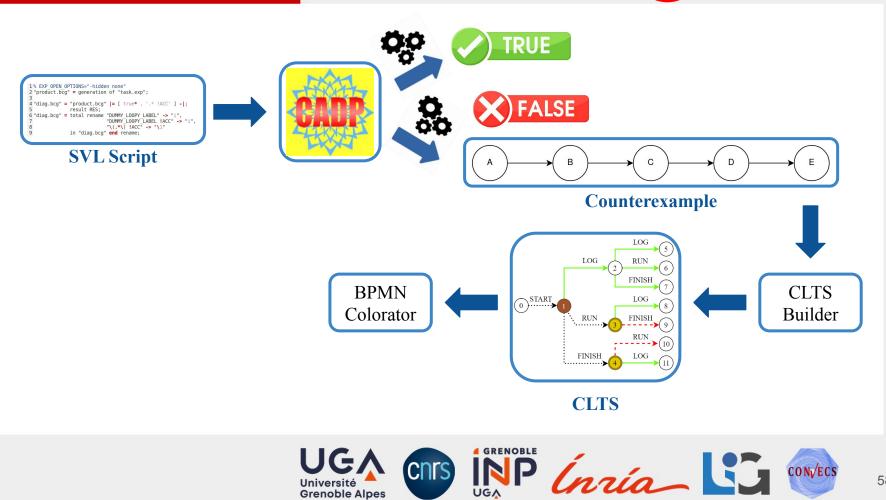
Counterexample



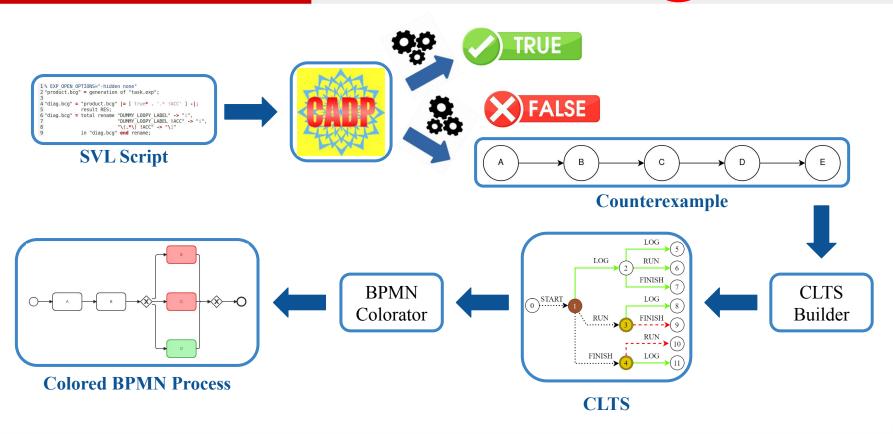




Grenoble Alpes

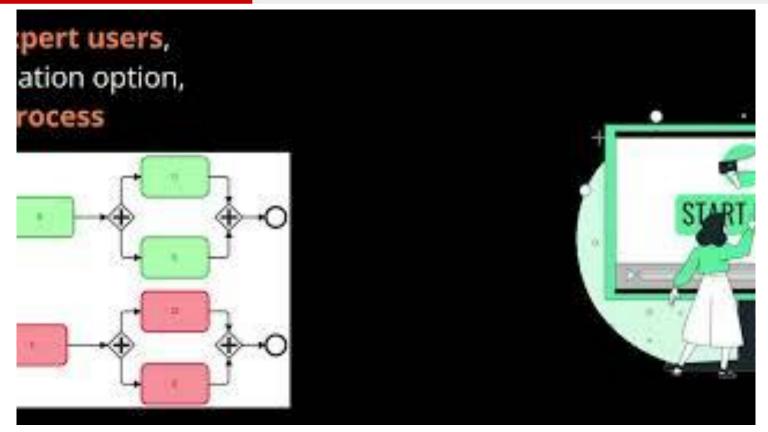


Université **Grenoble Alpes**



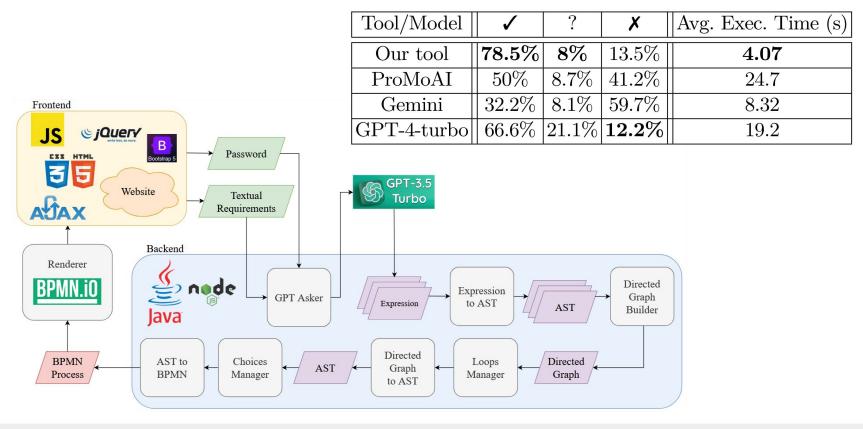


Tool – Presentation & Experiments





Tool – Presentation & Experiments





The LTL property is **correctly generated in 100%**. Necessary otherwise the verification makes no sense.

BPMN Process	States	Trans.	BPMN Gen. Time	Prop. Gen. Time	Model Check. Time	CLTS Cons. Time	BPMN Colo. Time
Evisa App. [19]	30	31	1.67s	1.43s	4.35s	3.12s	613ms
Patient Diag. [3]	38	40	2.71s	1.14s	4.57s	3.29s	661ms
Employee Rec. [9]	39	40	1.89s	1.33s	4.57s	3.22s	830ms
Employee Hiring [6]	78	105	2.17s	2.86s	4.34s	3.22s	758ms
Perish. Goods Trans. [20]	108	150	2.22s	1.1s	4.78s	3.31s	978ms
Acc. Open. Proc. [17]	304	657	2.31s	1.23s	4.56s	4.23s	1.42s
Hard. Ret. Ship. Proc. [10]	373	819	2.56s	1.1s	4.53s	3.13s	764ms
Online Shipping [16]	375	765	2.78s	1.07s	5.12s	3.27s	1.7s
Handcrafted 1	279k	1.63m	3.25s	1.07s	8s	14.6s	17.2s
Handcrafted 2	1.67m	11m	1.79s	1.27s	23.9s	5.42m	24m
Handcrafted 3	10m	75m	1.67s	1.18s	3.05m	>1h	>1h
Handcrafted 4	60m	503m	2.08s	867ms	27.7m	>1h	>1h
Handcrafted 5	362m	3.32b	2.31s	1.85s	>1h	>1h	>1h



The LTL property is **correctly generated in 100%**. Necessary otherwise the verification makes no sense.

BPMN Process	States	Trans.	BPMN Gen. Time	Prop. Gen. Time	Model Check. Time	CLTS Cons. Time	BPMN Colo. Time
Evisa App. [19]	30	31	1.67s	1.43s	4.35s	3.12s	613ms
Patient Diag. [3]	38	40	2.71s	1.14s	4.57s	3.29s	661ms
Employee Rec. [9]	39	40	1.89s	1.33s	4.57s	3.22s	830ms
Employee Hiring [6]	78	105	2.17s	2.86s	4.34s	3.22s	758ms
Perish. Goods Trans. [20]	108	150	2.22s	1.1s	4.78s	3.31s	978ms
Acc. Open. Proc. [17]	304	657	2.31s	1.23s	4.56s	4.23s	1.42s
Hard. Ret. Ship. Proc. [10]	373	819	2.56s	1.1s	4.53s	3.13s	764ms
Online Shipping [16]	375	765	2.78s	1.07s	5.12s	3.27s	1.7s
Handcrafted 1	279k	1.63m	3.25s	1.07s	<u>8s</u>	14.6s	17.2s
Handcrafted 2	1.67m	11m	1.79s	1.27s	23.9s	5.42m	24m
Handcrafted 3	10m	75m	1.67s	1.18s	3.05m	>1h	>1h
Handcrafted 4	60m	503m	2.08s	867ms	27.7m	>1h	>1h
Handcrafted 5	362m	3.32b	2.31s	1.85s	>1h	>1h	>1h

Starts getting (very) long on LTSs with millions of states

Inría 13





CON/ECS

Conclusion

In this work, we proposed a **9 steps approach** aiming at automatically designing **syntactically and semantically correct BPMN** processes from a **textual description** of the requirements.

The main **perspectives** of this work are:

- Support unbalanced workflows
- Make use of more recent versions of GPT
- Diversify supported LTL properties
- Enrich the training dataset
- Enlarge the supported BPMN syntax

