Generative Artificial Intelligence in Chemical Process Engineering

Artur M. Schweidtmann

Graph Seminar

Process Intelligence Research
Dept. Chemical Engineering
Delft University of Technology

3. October 2024

TUDelft

Process Intelligence RESEARCH

Generative AI led to breakthroughs in multiple domains

- e.g.,

  - Natural language[1]: ChatGPT

  - Images[2]: DALL·E 3

  - Video: Sora

  - Molecules[3]: MoLeR

- Generative AI in process systems engineering

  - Is it feasible?

  - How would it look like?

  - What does it take to develop it?
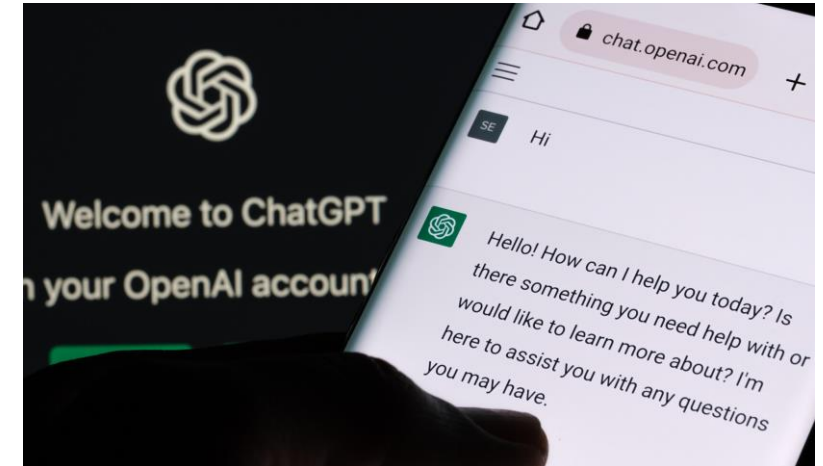
[1] OpenAI (2023). GPT-4 Technical Report. arXiv preprint arXiv: 2303.08774.
[2] Shi, Z., Zhou, X., Qiu, X., & Zhu, X. (2020). Improving image captioning with better use of captions. arXiv preprint arXiv:2006.11807.
[3] Maziarz, K., Jackson-Flux, H., Cameron, P., Sirockin, F., Schneider, N., Stiefl, N., ... & Brockschmidt, M. (2021). Learning to extend molecular scaffolds with structural motifs. arXiv preprint arXiv:2103.03864.
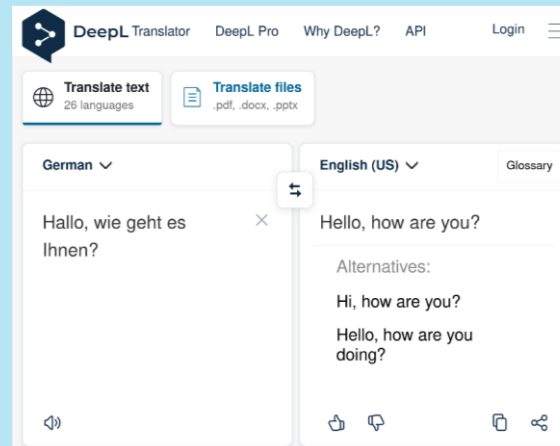
Three exciting generative AI applications

Auto-completion of code
(e.g., GitHub Copilot[1])

Machine translation
(e.g., DeepL[2])

Auto-correction of text
(e.g., Grammarly[3])

[1] https://github.com/features/copilot [2] https://www.deepl.com/ [3] https://app.grammarly.com/

Schweidtmann Graph Seminar

Piping and instrumentation diagrams

...and their counterpart in process engineering

Auto-completion of code
(e.g., GitHub Copilot)

Machine translation
(e.g., DeepL)
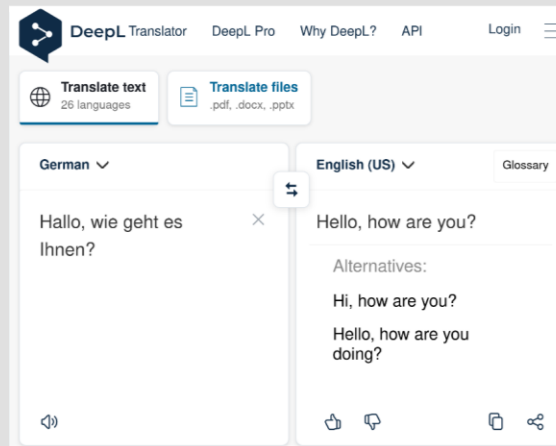
Auto-correction of text
(e.g., Grammarly)

Auto-completion of flowsheets

…and their counterpart in process engineering

Auto-completion of flowsheets

Machine translation
(e.g., DeepL)

Auto-correction of text
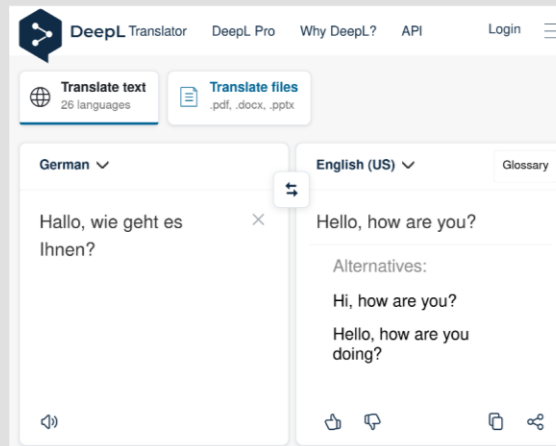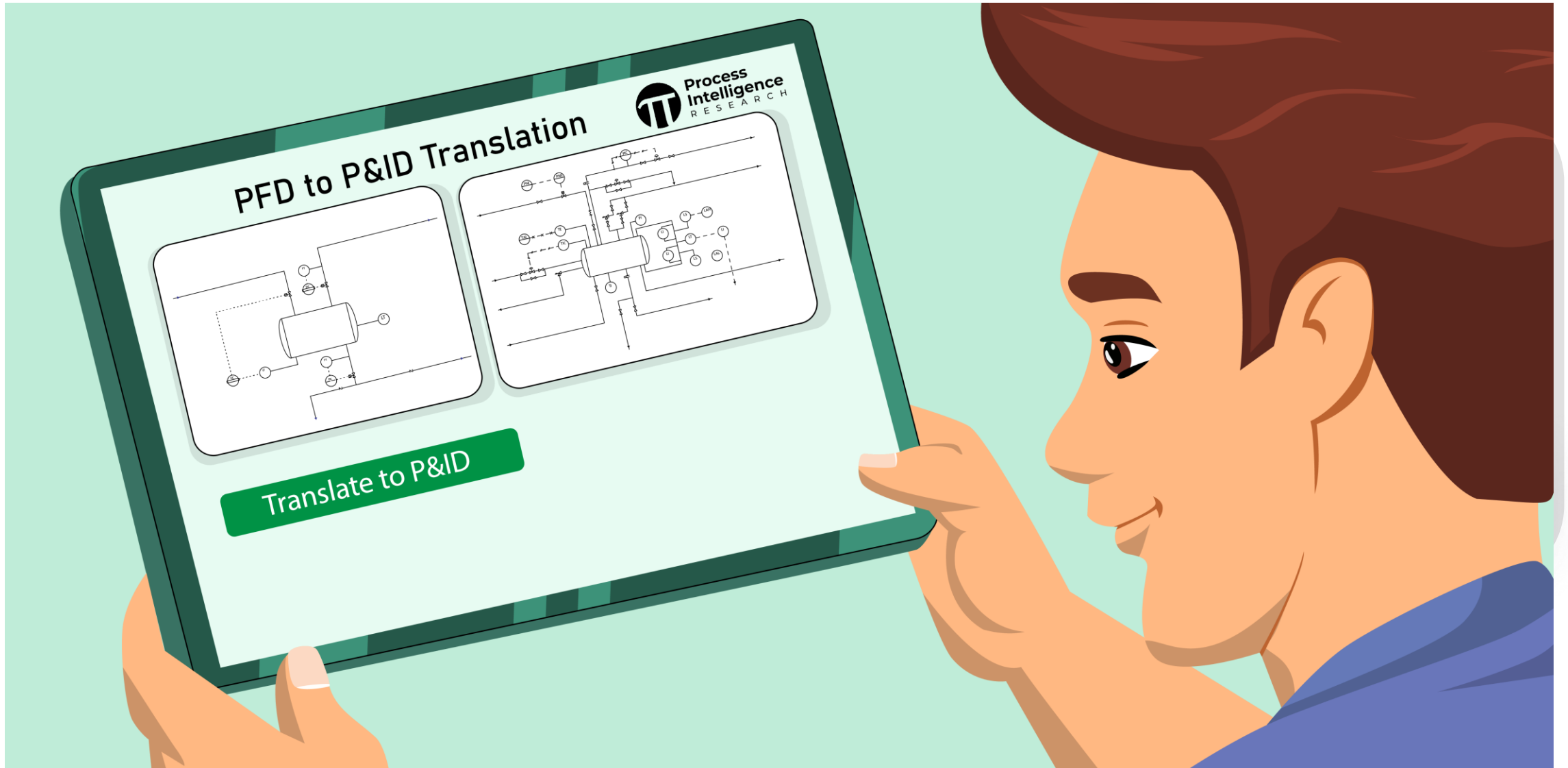*(e.g., Grammarly)*

[1] Schweidtmann, A. M. (2024). Generative artificial intelligence in chemical engineering. Nature Chemical Engineering.

PFD to P&ID translation

…and their counterpart in process engineering

Auto-completion of flowsheets

PFD to P&ID translation

Auto-correction of text
(e.g., Grammarly)

[1] Schweidtmann, A. M. (2024). Generative artificial intelligence in chemical engineering. Nature Chemical Engineering.

Auto-correction of P&IDs
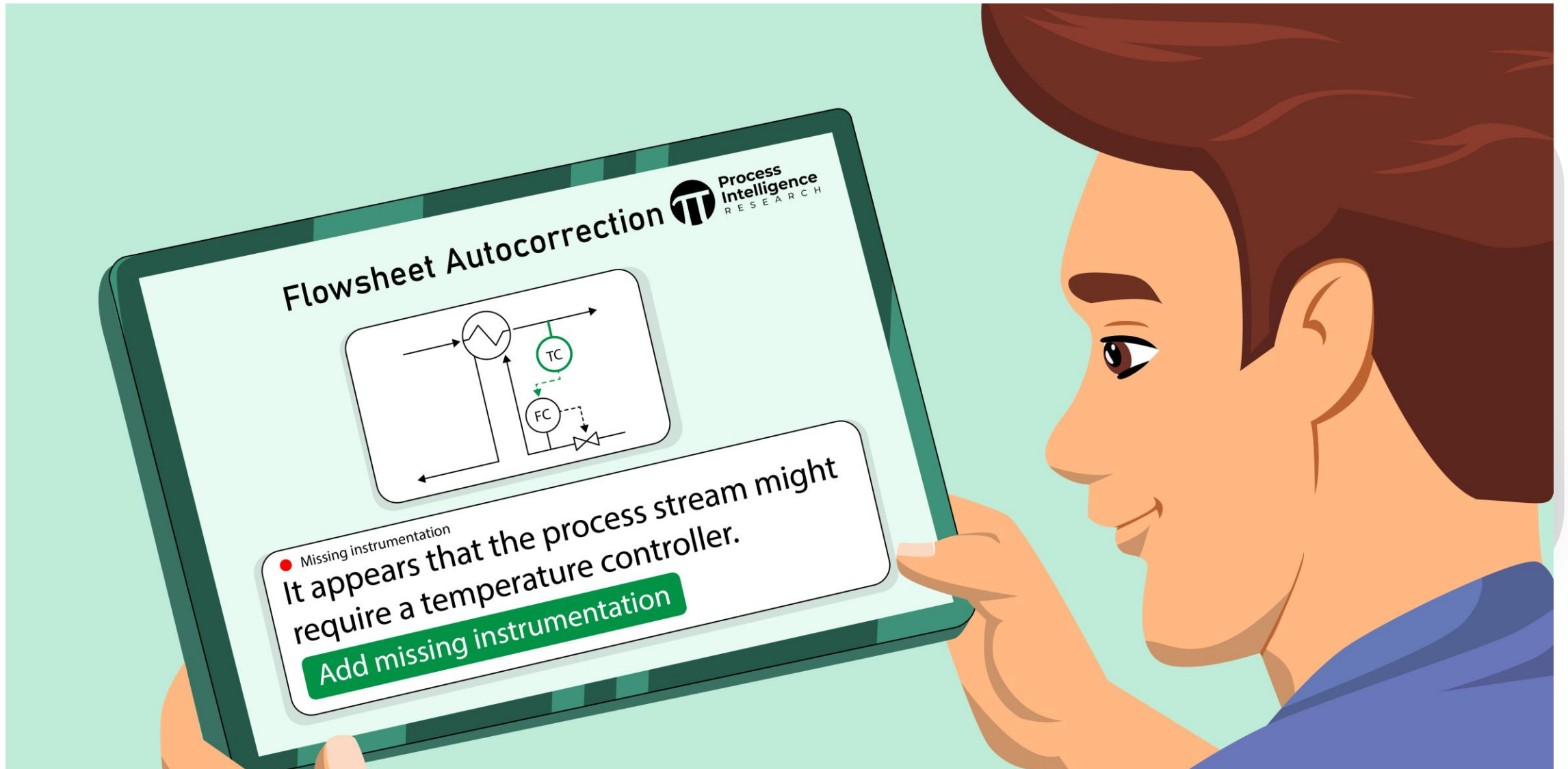
Our case studies for Generative AI in PSE

Auto-completion of flowsheets
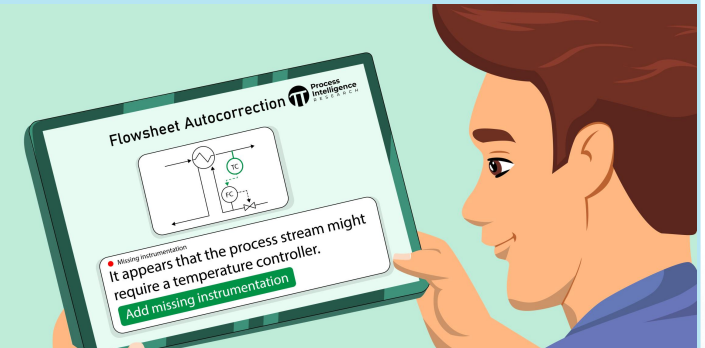
PFD to P&ID translation

Auto-correction of P&IDs

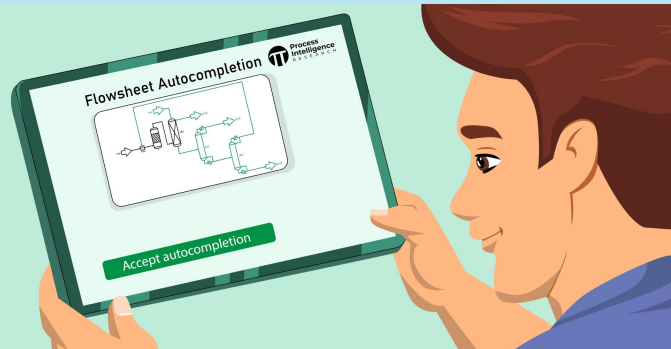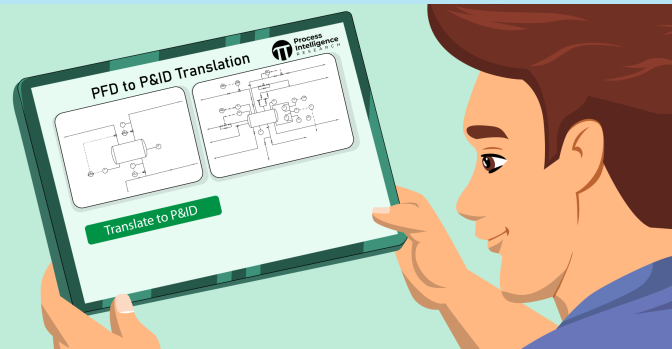[1] Schweidtmann, A. M. (2024). Generative artificial intelligence in chemical engineering. Nature Chemical Engineering.

What does it take to develop generative AI algorithms?

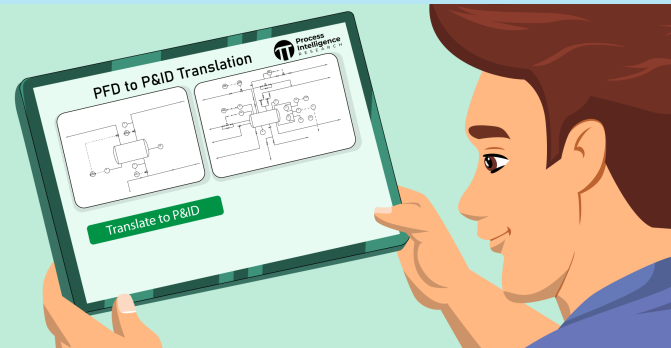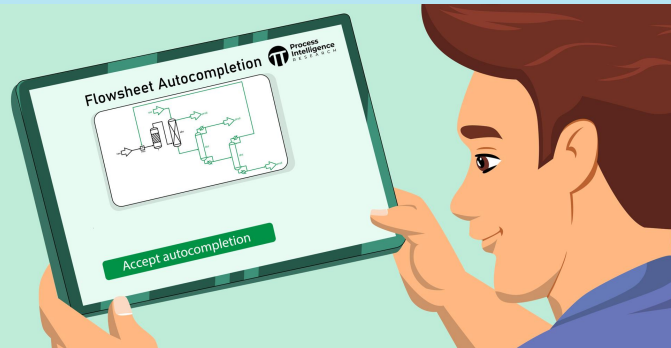|	Big data	Information representation	Model architecture	Mechanistic knowledge
ChatGPT	57 TB	Text	Transformer	
Process engineering	?	?	?	?

What does it take to develop generative AI algorithms?

Big data	Information representation	Model architecture	Mechanistic knowledge

Process engineering
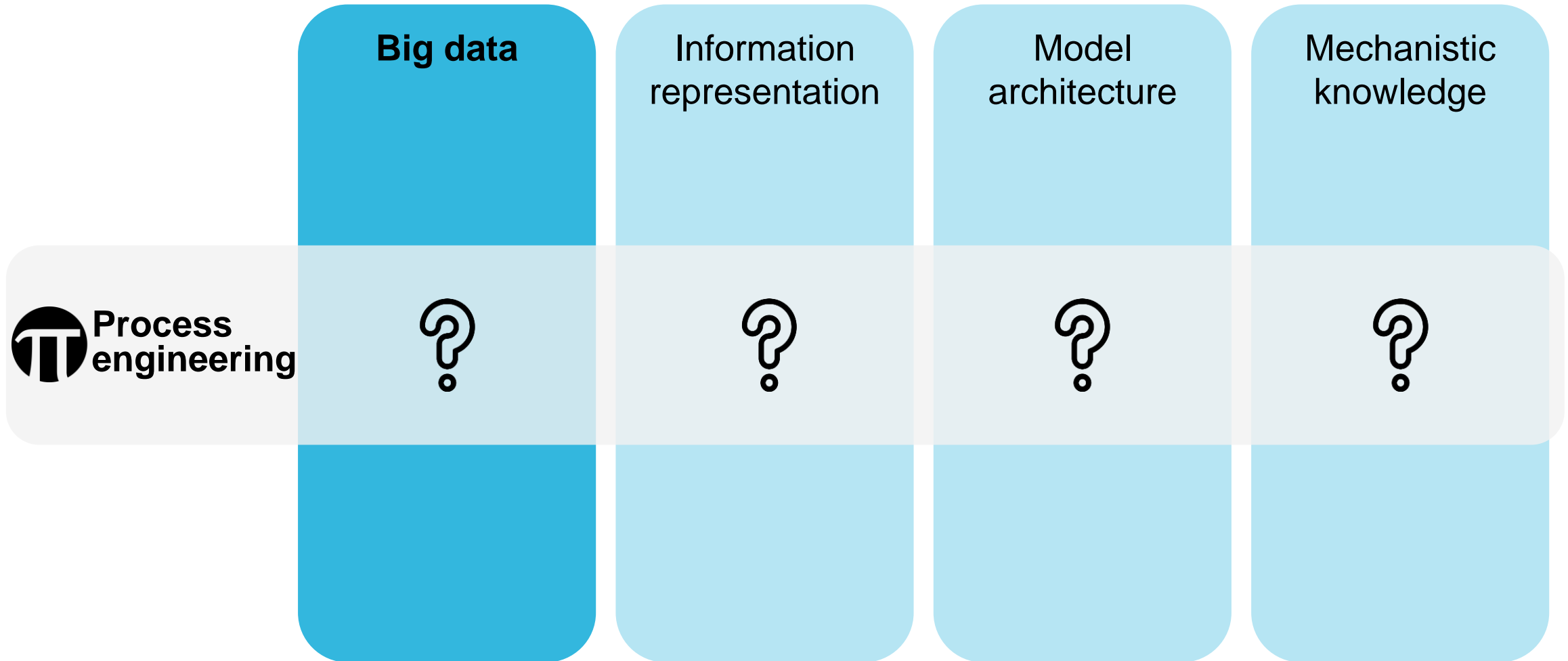
Is chemical engineering a big data domain?

- Our domain is not a *true* big data domain[1,2]

  - Limited data domain

  - Heterogeneous

  - Not findable

  - Not machine readable

- Flowsheets are *hidden* in

  - Company reports
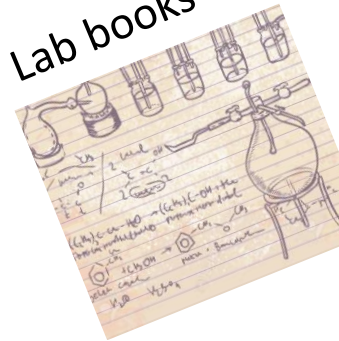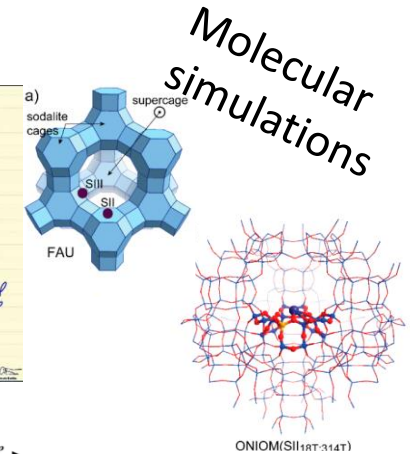
  - Books, scientific publications, and patents

Lab books

Property data

Molecular simulations

FAU

ONIOM(SII_{1RT-314T})
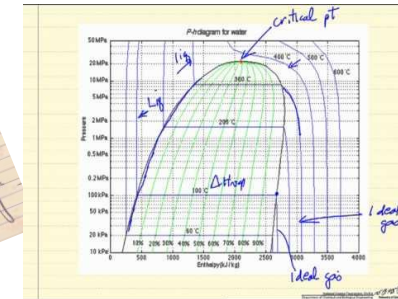
Flowsheets

Simulation files

Publications

[1] Venkatasubramanian, V. (2019). The promise of artificial intelligence in chemical engineering: Is it here, finally?. AIChE Journal, 65
[2] Schweidtmann, A. M., Esche, E., Fischer, A., Kloft, M., Repke, J. U., Sager, S., & Mitsos, A. (2021). Machine learning in chemical engineering: A perspective. Chemie Ingenieur Technik, 93(12), 2029-2039.

Flowsheet mining from publications and patents
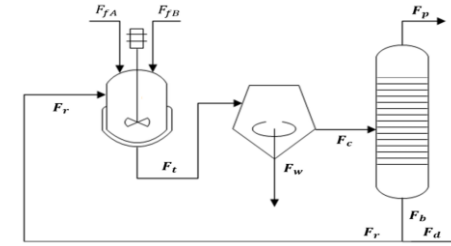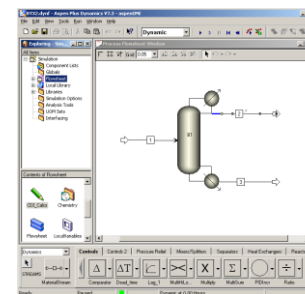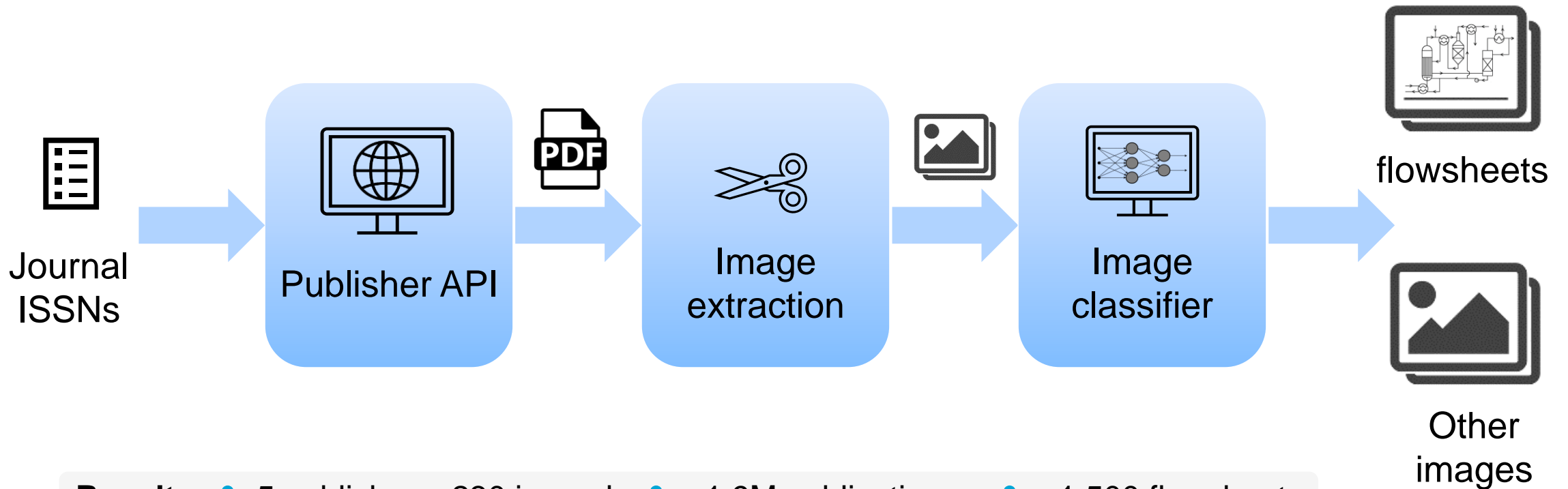

Results: • 5 publishers, 230 journals • >1.9M publications • >1,500 flowsheets

[1] Balhorn, L. S., Gao, Q., Goldstein, D., & Schweidtmann, A. M. (2022). Flowsheet recognition using deep convolutional neural networks. In Computer Aided Chemical Engineering (Vol. 49, pp. 1567-1572). Elsevier.

DigiCo – Digitization Companion
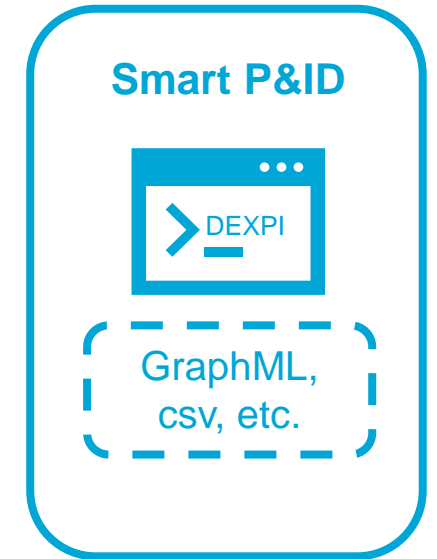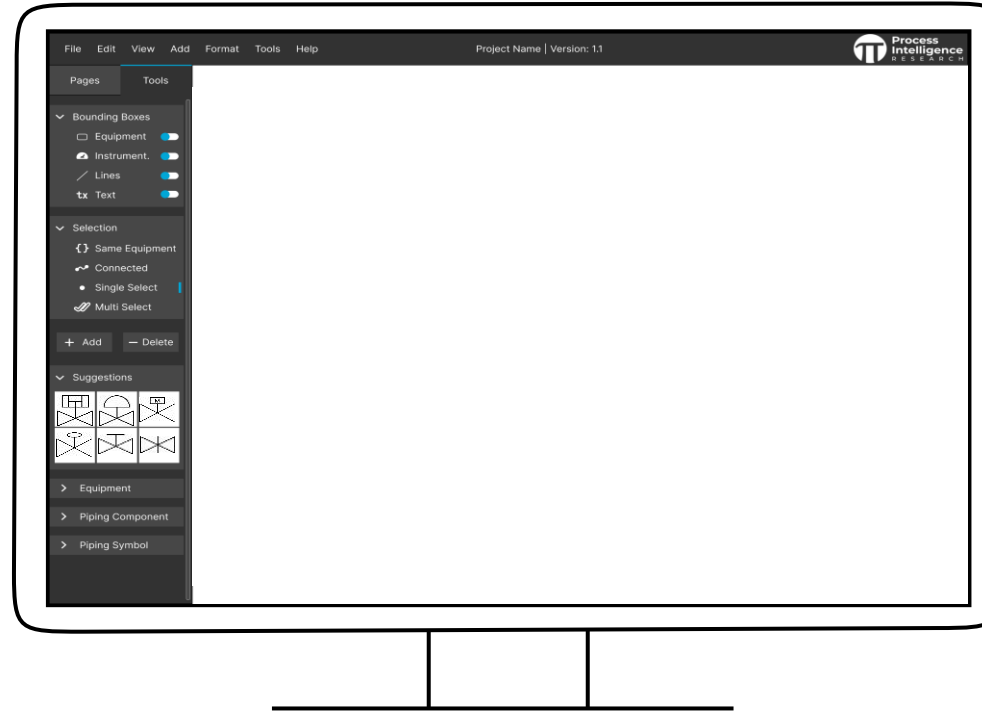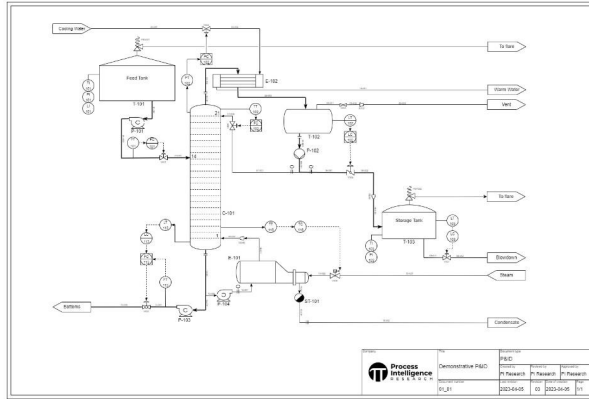
[1] Theisen, M. F., Flores, K. N., Balhorn, L. S., & Schweidtmann, A. M. (2023). Digitization of chemical process flow diagrams using deep convolutional neural networks. Digital Chemical Engineering, 6, 100072.

What does it take to develop generative AI algorithms?

	Big data	Information representation	Model architecture	Mechanistic knowledge
Process engineering	Digital flowsheets	?	?	?
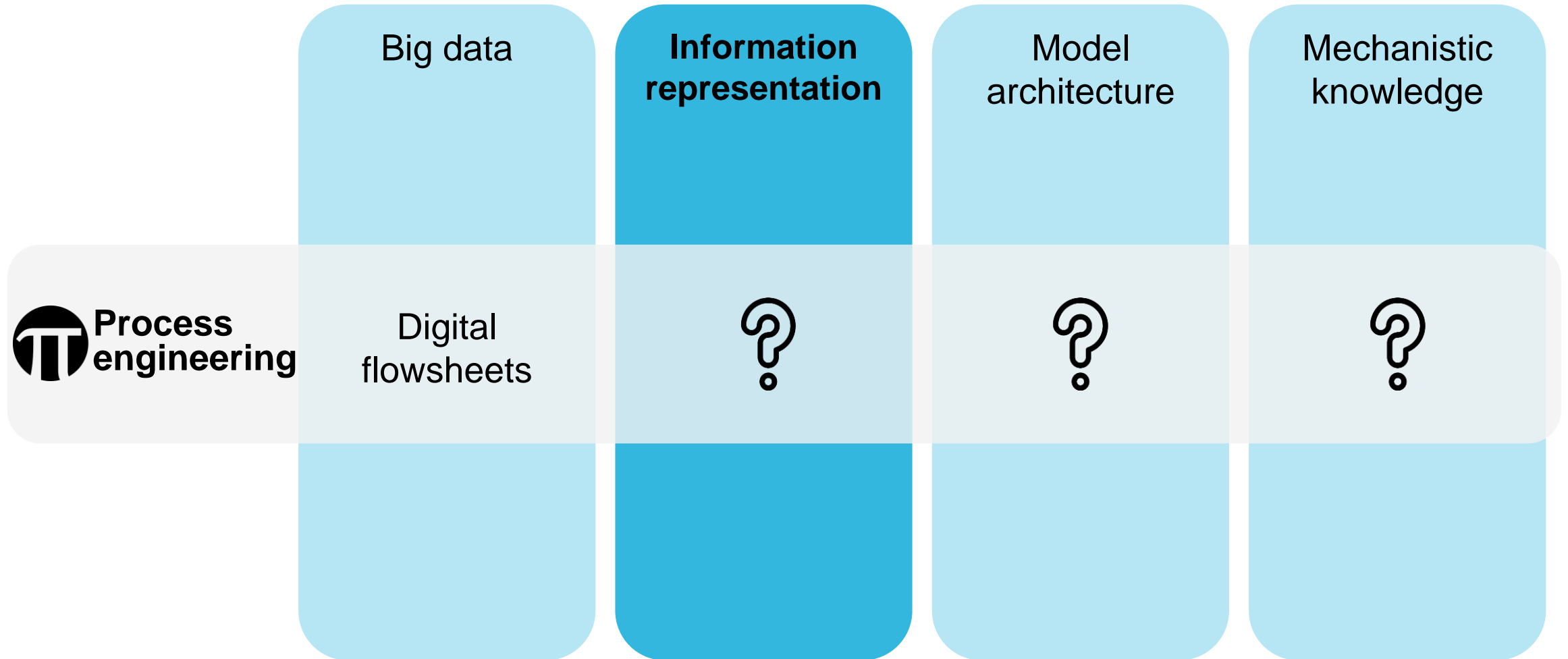
Information representation of flowsheets
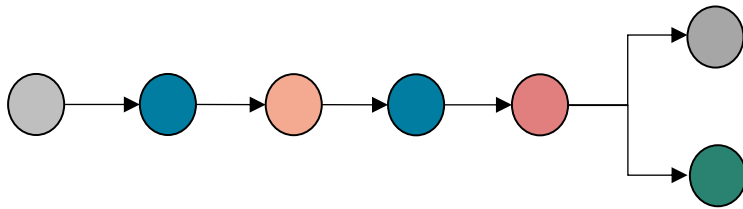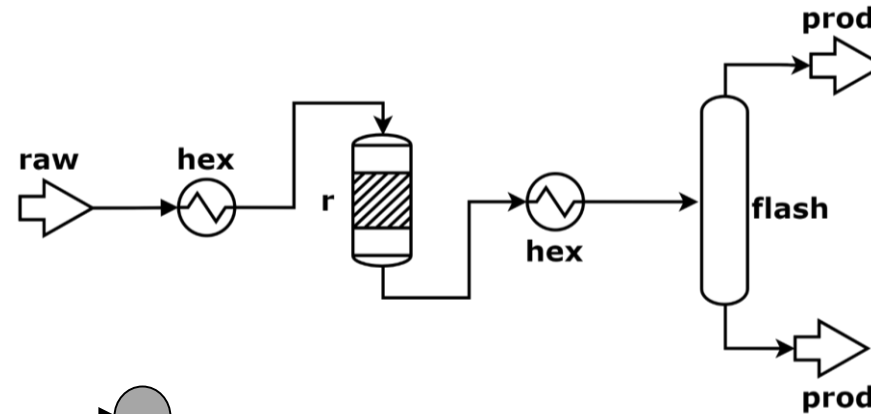
(raw)(hex)(r)(hex) (flash)[{tout}(prod)]{bout}(prod)

Flowsheet graphs[1] $G = (V, E)$

- Unit operations as nodes and streams as edges

- Add. information as feature vectors

- Variations incl. direct, hyper-, knowledge graphs (e.g., based on DEXPI)

SFILES[2-4]

- Unique text representation of flowsheet topology

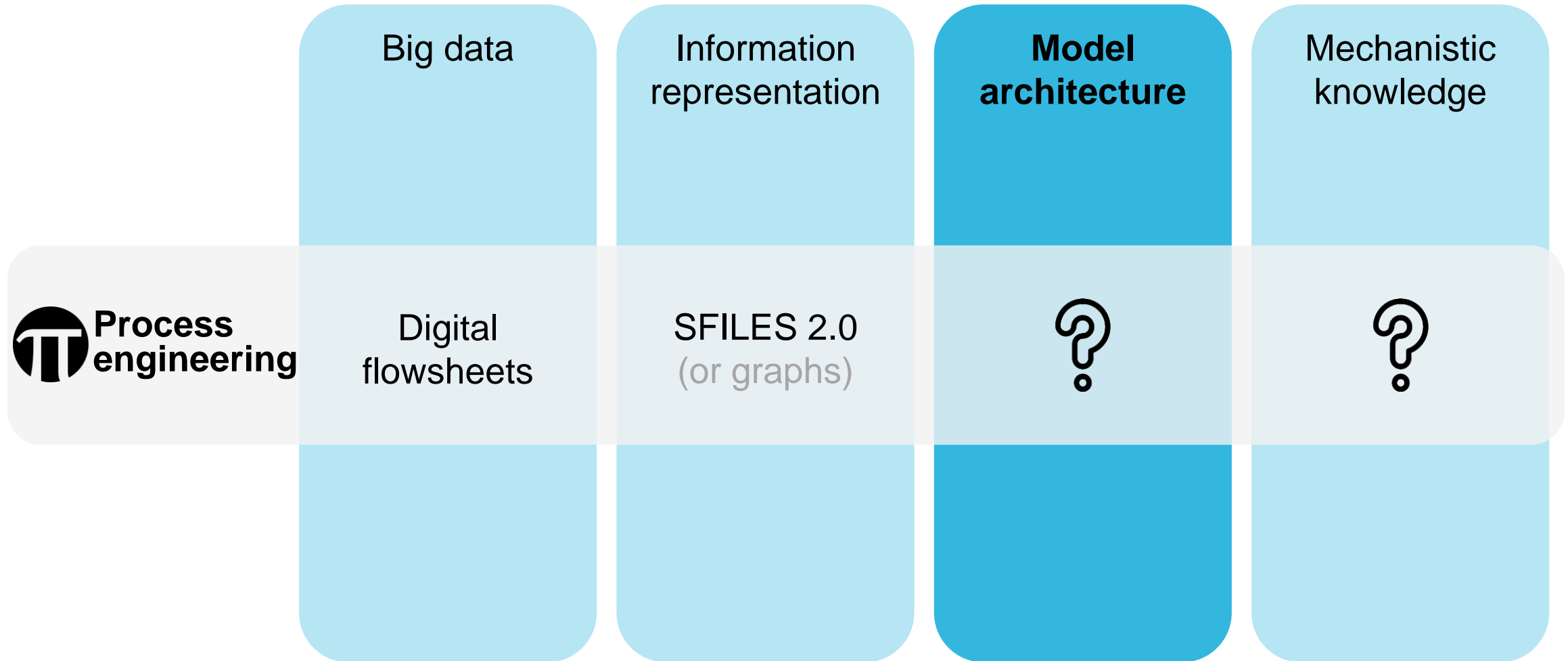- Inspired by SMILES for molecules

- Currently limited to topology

[1] Stops, L., Leenhouts, R., Gao, Q., & Schweidtmann, A. M. (2023). Flowsheet generation through hierarchical reinforcement learning and graph neural networks. AIChE Journal, 69(1), e17938.
[2] d'Anterroches, L. (2005). *Process Flowsheet Generation & Design through a Group Contribution Approach.* [CAPEC], Department of Chemical Engineering, Technical University of Denmark.
[3] Vogel, G., Hirtreiter, E., Schulze Balhorn, L., & Schweidtmann, A. M. (2023). SFILES 2.0: an extended text-based flowsheet representation. *Optimization and Engineering*, 1-23.
[4] Mann, V., Gani, R., Venkatasubramanian, V. (2023). Intelligent Process Flowsheet Synthesis and Design using Extended SFILES Representation. *ESCAPE33*

# What does it take to develop generative AI algorithms?

	Big data	Information representation	**Model architecture**	Mechanistic knowledge
Process engineering	Digital flowsheets	SFILES 2.0 (or graphs)	?	?

Auto-completion of flowsheets

Proof of concept
Limited to topology information, training bias, potential errors in prediction...

raw

raw

mix

r

abs

prod

dist

prod

dist

prod

prod

Absorbent could be recovered
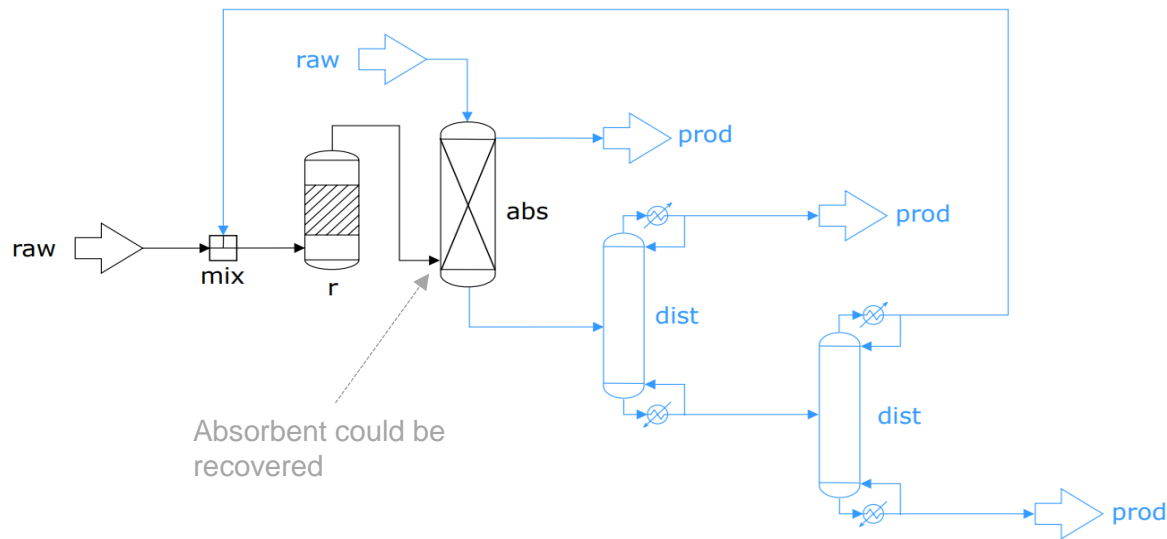
Dataset

- 223 real and 8k synthetic flowsheets
- 53 vocabulary size

Model

- GPT-2 using decoder-only architecture (85.9M param)
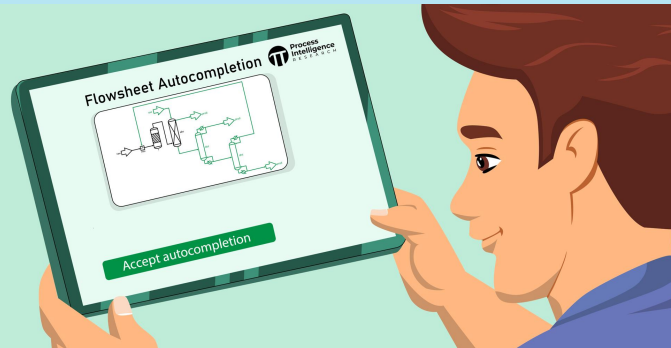- **In**: Incomplete PFD; **Out**: Next unit

Results: Pretraining is beneficial

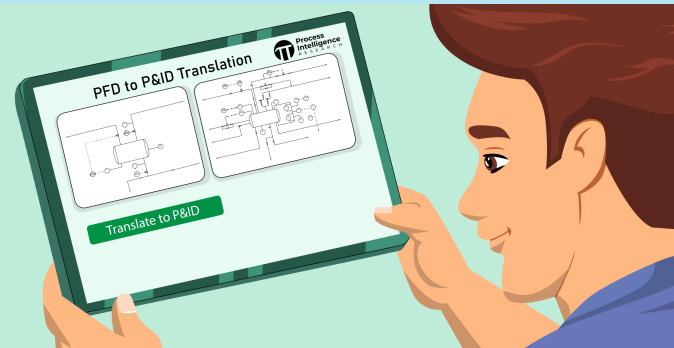- Perplexity ~25.91 *w/o* pretraining
- Perplexity ~4.75 *w/* pretraining

[1] Vogel, G., Schulze Balhorn, L., & Schweidtmann, A. M., (2023), Learning from flowsheets: A generative transformer model for autocompletion of flowsheets, Computers & Chemical Engineering
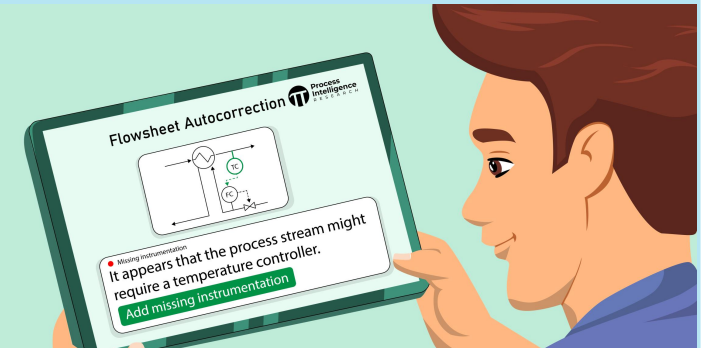
Case studies

Auto-completion of flowsheets

PFD to P&ID translation

Auto-correction of P&IDs

[1] Schweidtmann, A. M. (2024). Generative artificial intelligence in chemical engineering. Nature Chemical Engineering.
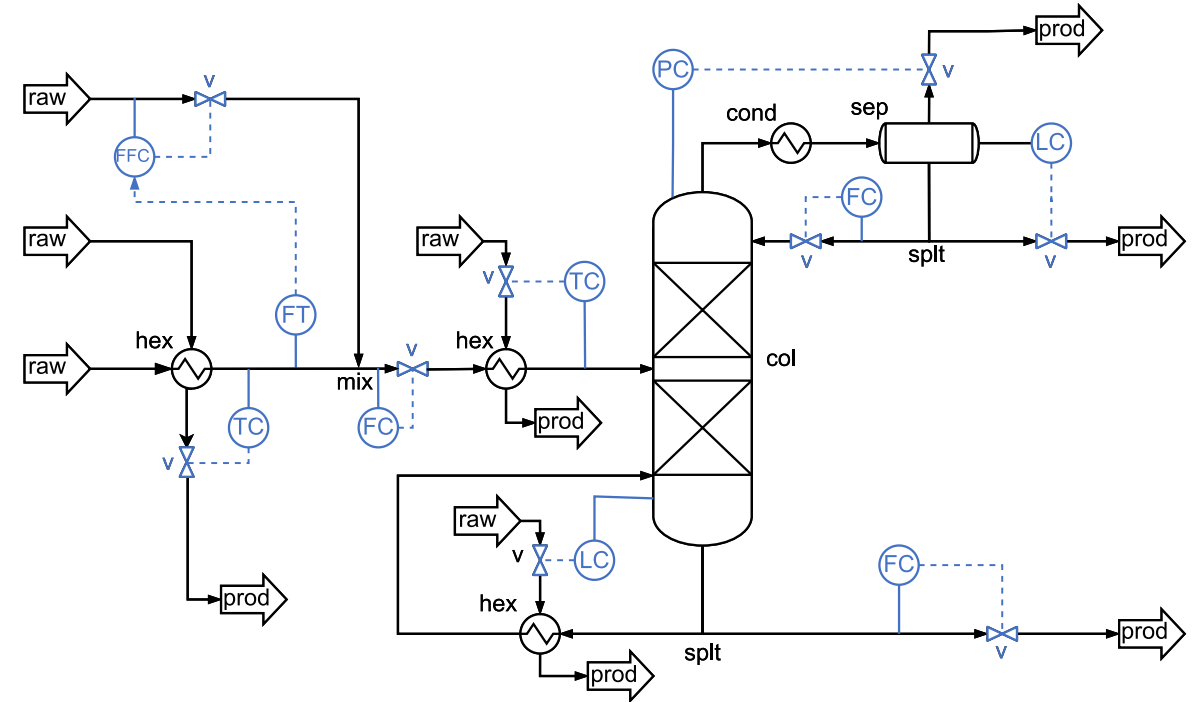
Control structure prediction

Dataset

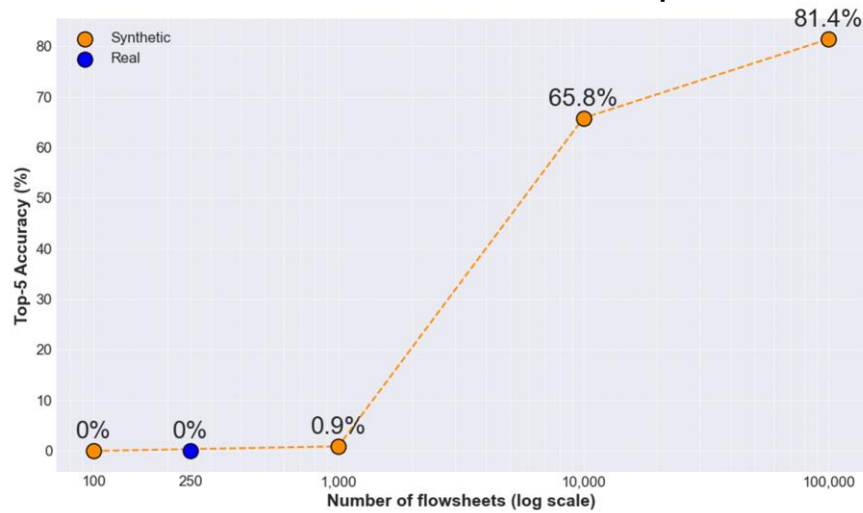- 100,000 synthetic flowsheets with control structure
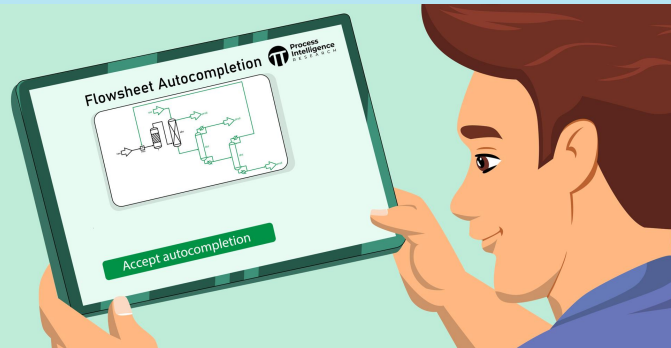- 250 real flowsheets with control structure

Model

- T5 encoder-decoder model (7.9M param)
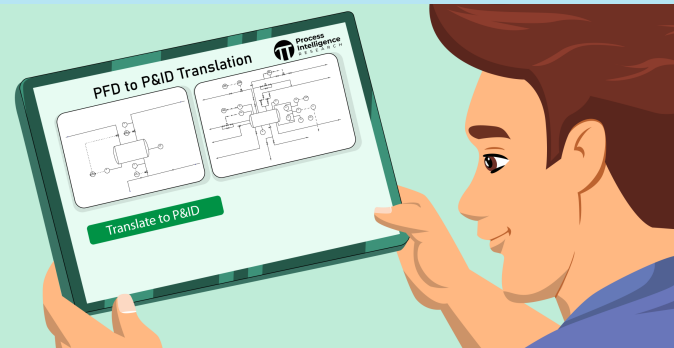- **In**: PFD; **Out**: PFD w/ control loops

[1] Hirtreiter, E., Schulze Balhorn, L., & Schweidtmann, A. M. (2023). Toward automatic generation of control structures for process flow diagrams with large language models. AIChE Journal
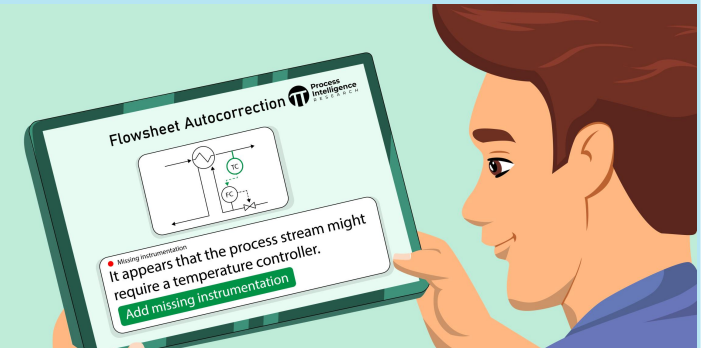
Case studies

Auto-completion of flowsheets

PFD to P&ID translation

Auto-correction of P&IDs

[1] Schweidtmann, A. M. (2024). Generative artificial intelligence in chemical engineering. Nature Chemical Engineering.
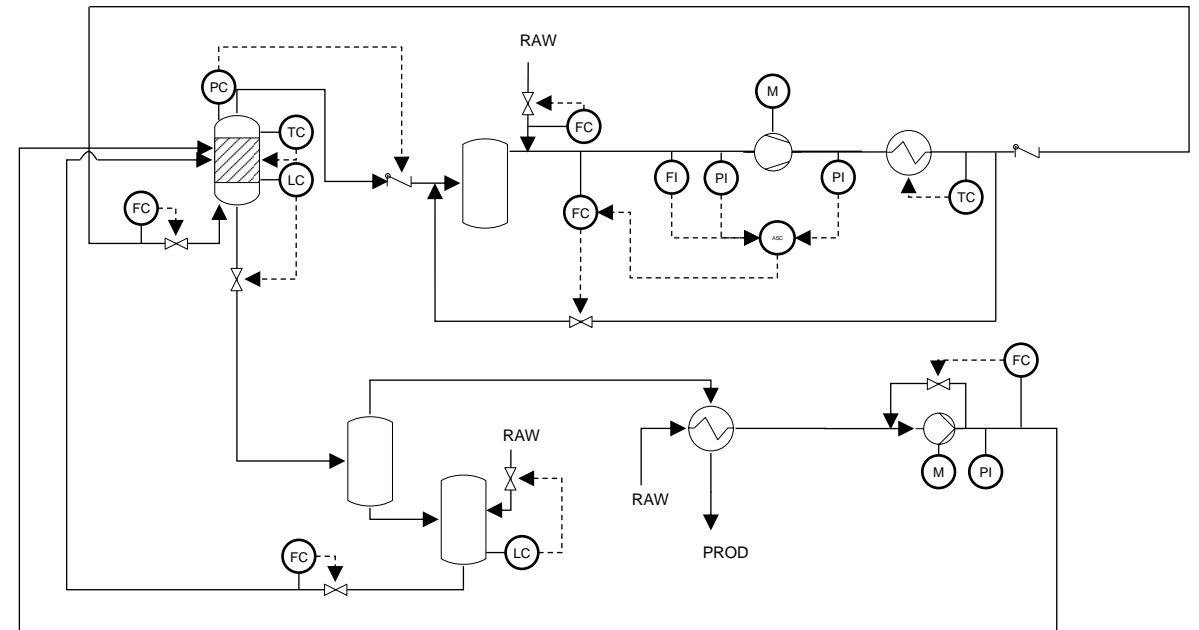
Auto-correction of flowsheets

Dataset

- 500k synthetic flowsheets
- 53 vocabulary size

Model

- T5 encoder-decoder (7.9M param)
- **In**: PFD w/ errors; **Out**: Corrected PFD

Results

- Top-1 accuracy 82.10%
- Top-5 accuracy 84.40%

[1] Balhorn, L. S., Caballero, M., & Schweidtmann, A. M. (2023). Toward autocorrection of chemical process flowsheets using large language models. arXiv preprint arXiv:2312.02873.
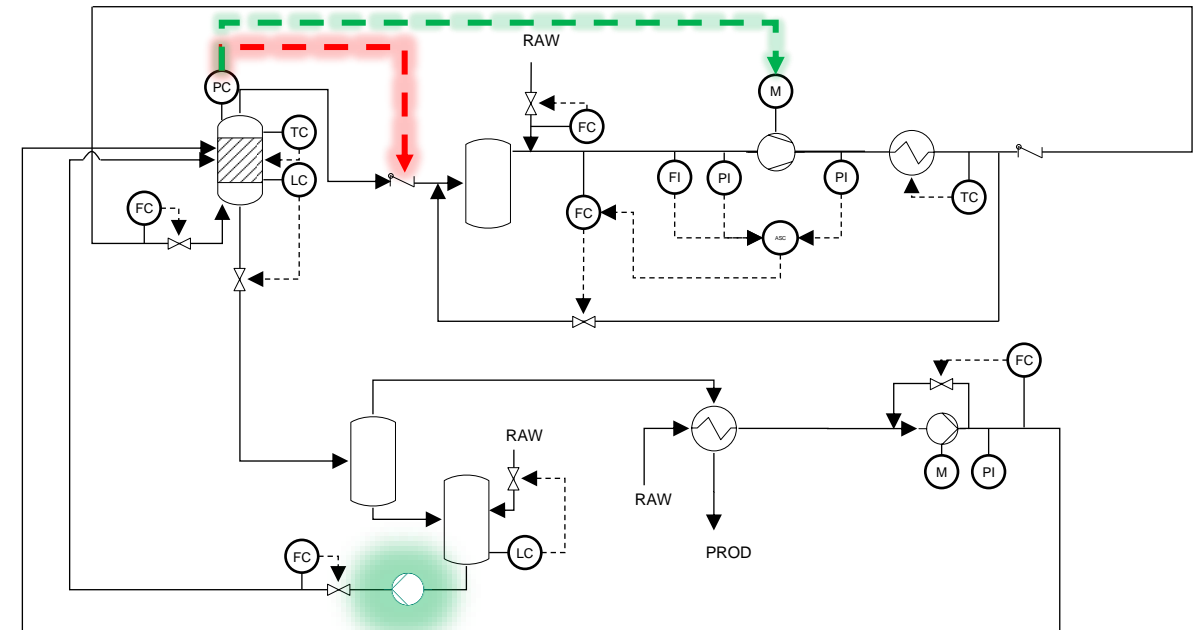
Auto-correction of flowsheets

Dataset

- 500k synthetic flowsheets
- 53 vocabulary size

Model

- T5 encoder-decoder (7.9M param)
- **In**: PFD w/ errors; **Out**: Corrected PFD
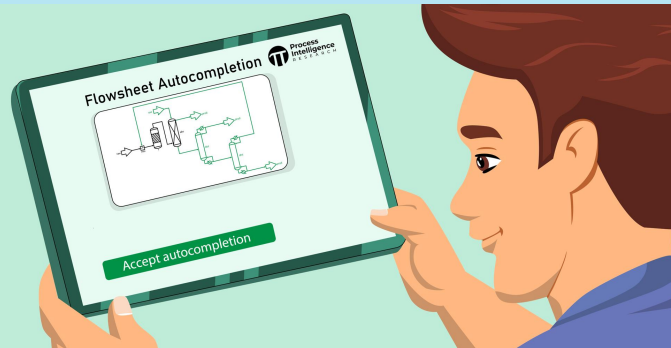
Results
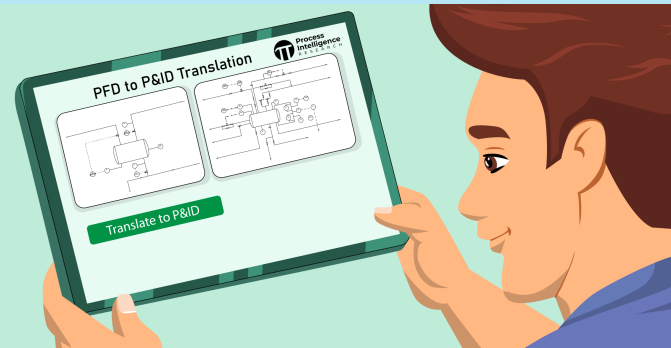
- Top-1 accuracy 82.10%
- Top-5 accuracy 84.40%

[1] Balhorn, L. S., Caballero, M., & Schweidtmann, A. M. (2023). Toward autocorrection of chemical process flowsheets using large language models. arXiv preprint arXiv:2312.02873.

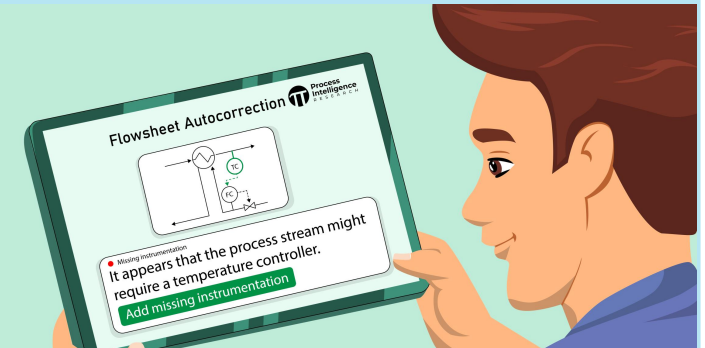# PILOT - Process Intelligence Learning and Optimization Toolbox

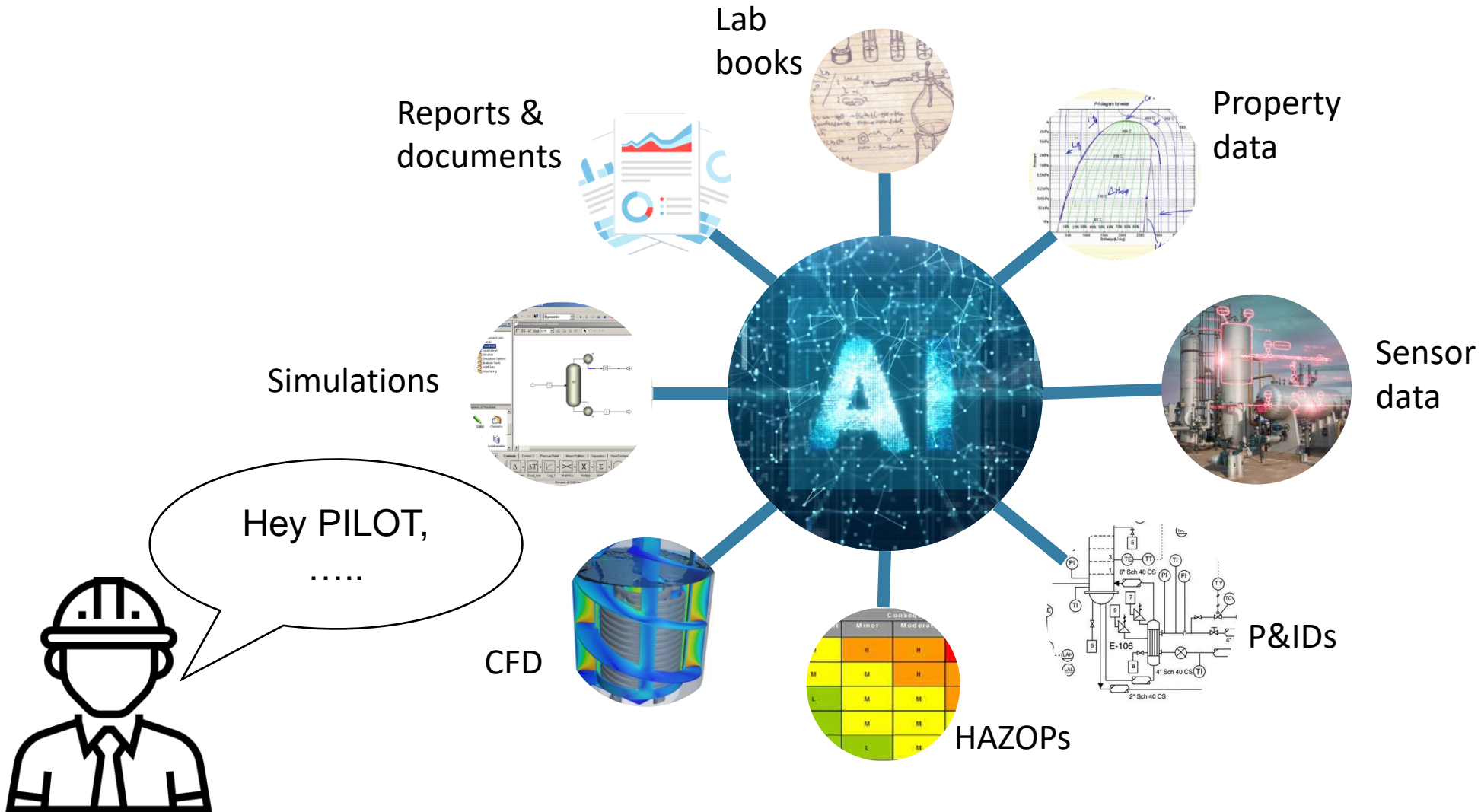**Auto-completion of flowsheets**

PFD to P&ID translation
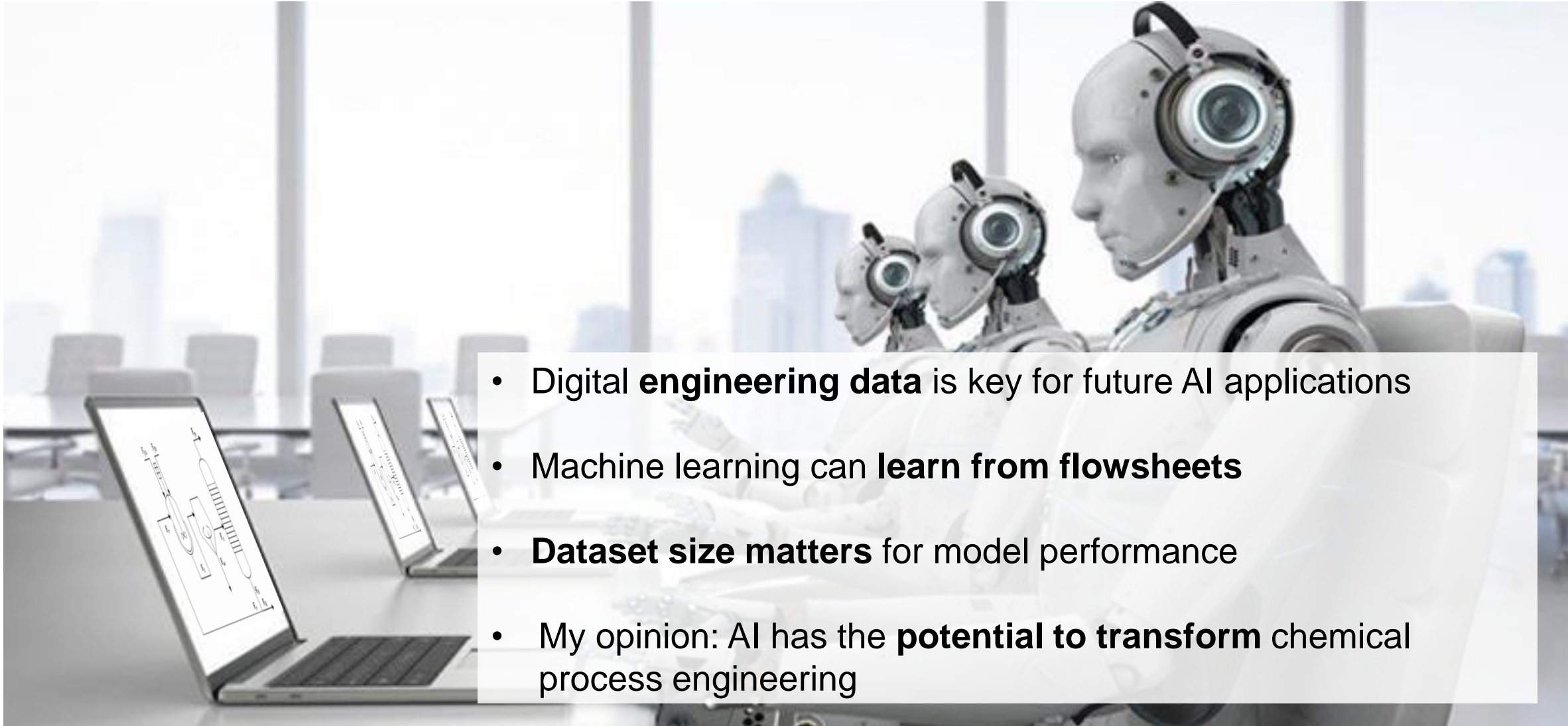
Auto-correction of P&IDs

[1] Schweidtmann, A. M. (2024). Generative artificial intelligence in chemical engineering. Nature Chemical Engineering.

PILOT will integrate multi-modal information

Lab books

Reports & documents

Property data

Simulations

Sensor data
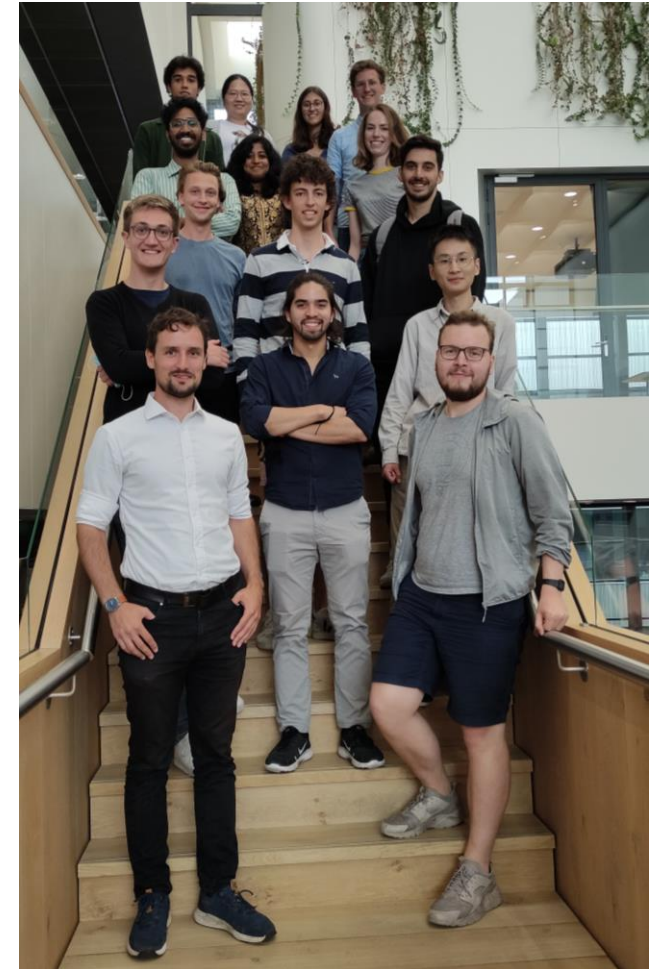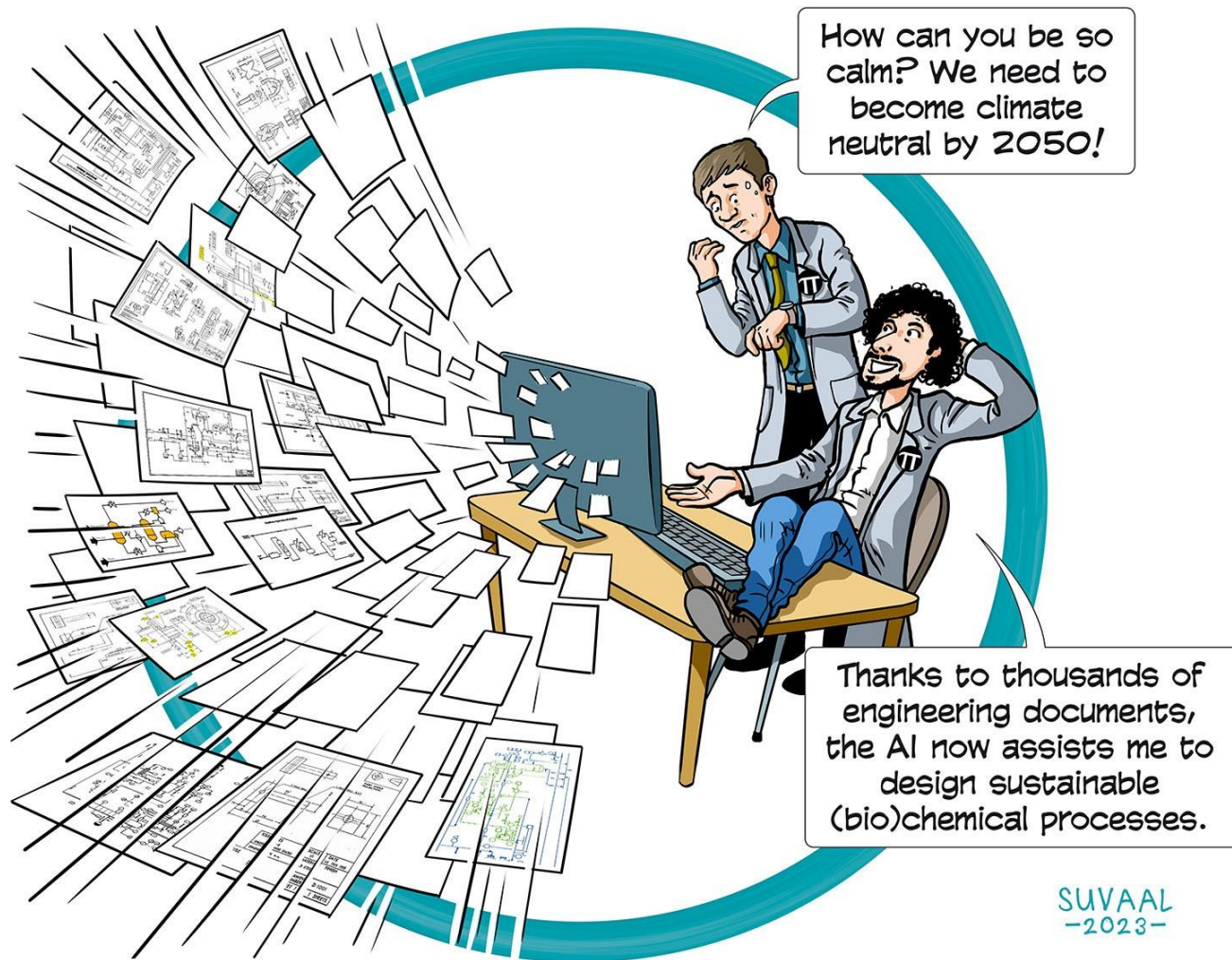
Hey PILOT, …..

CFD

HAZOPs

P&IDs

[1] Schweidtmann, A. M. (2024). Generative artificial intelligence in chemical engineering. Nature Chemical Engineering.

Conclusions

- Digital **engineering data** is key for future AI applications

- Machine learning can **learn from flowsheets**

- **Dataset size matters** for model performance

- My opinion: AI has the **potential to transform** chemical process engineering

Thanks to the amazing team

Thank you very much for your attention!

✉ a.schweidtmann@tudelft.nl

🏠 https://www.pi-research.org/

𝕏 https://twitter.com/ASchweidtmann

in https://www.linkedin.com/in/schweidtmann/

▶ https://www.youtube.com/@ProcessIntelligenceResearch