2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC

Host Sponsor UNC Charlotte

Paper ID #50547

Effective Use of Generative Artificial Intelligence in Chemical Process Design Undergraduate Courses

Dr. Daniela Galatro, University of Toronto
Dr. Sourojeet Chakraborty, Johns Hopkins University

Dr. Sourojeet Chakraborty is a Lecturer at the Department of Chemical & Biomolecular Engineering at Johns Hopkins University, USA, having earned his Ph.D. (University of Toronto) and MSc. & DIC (Imperial College London) in chemical engineering. He was a Postdoctoral Fellow at the Department of Earth Sciences, and a Junior Fellow of the Massey College, University of Toronto. His research interests lie in transport phenomena, fluid mechanics, and the design of complex fluids. Within engineering education, his research interests lie in mapping Industry 5.0 to Education 5.0, curriculum design, pedagogical strategies for innovation, enriching and empowering the student learning experience, and promoting Diversity, Equity, and Inclusion (DEI) within the university landscape.

Effective Use of Generative Artificial Intelligence in Chemical Process Design Undergraduate Courses

Daniela Galatro[1] and Sourojeet Chakraborty[2]

[1]Department of Chemical Engineering & Applied Chemistry, University of Toronto ON M5S 3E5 Canada
[2]Department of Chemical & Biomolecular Engineering, Johns Hopkins University, Baltimore MD 21218 USA

Correspondence email: daniela.galatro@utoronto.ca

Abstract

Generative Artificial Intelligence (GenAI) has been integrated into chemical process design engineering education at universities. This integration includes applications in control structure topologies, the development and autocompletion of process diagrams, and the generation of process flow diagrams, which are cross-validated via process simulators, among others. While early evidence suggests promising learning outcomes and student engagement, evidence for its implementation effectiveness remains relatively limited. Most studies involve small cohorts and pilot-scale initiatives, which might not be representative enough to make sound conclusions. On the other hand, review papers highlight gaps in the research on GenAI in undergraduate engineering, as pedagogical frameworks for integrating these tools are still being conceptualized and implemented in engineering education. This work presents a systematic review of the use of GenAI in undergraduate chemical process design courses, focusing on clustering the specific domains of its use and on effectiveness assessment tools to meet learning outcomes and foster the mapping of Education 5.0 and Industry 5.0 elements. Moreover, we offer insights into a potential roadmap for GenAI integration into chemical engineering capstone design courses, covering the current development, deployment, assessment, and challenges of examples implemented in a third-year capstone course. In these examples, we explore strategies for "effective prompt engineering", design pipelines, and cross-validation of GenAI responses, which support the development of process deliverables, including process flow diagrams, piping and instrumentation diagrams, heat and mass balance, safety studies, circularity metrics, and economic evaluations of chemical processes. While this course focuses on team strategies for engineering design and the workflow is not entirely based on an open-ended project, it offers an earlier opportunity in our curriculum to guide students in using GenAI for chemical process design. This course is delivered in the semester right before the final capstone course in the fourth year of our program. Preliminary results reflect divergent levels of certainty and usefulness of GenAI tools among students when developing and/or validating process deliverables. Safety-related deliverables supported with Gen-AI are fairly described (according to the instructor and subject matter expert) because prompts are scaffolded with previous deliverables, including process diagrams and heat and material balance. However, process diagrams notably differ from standardized representations and, in some cases, lack well-structured control loops. Heat and material balances are efficiently prompted with process simulations, while deep troubleshooting is required to refine calculations. Effectiveness metrics for measuring GenAI prompt engineering and cross-validation are estimated and assessed by the subject matter expert. These metrics include comparing deliverables against codes, standards, and regulations, as well as reference-based calculations developed in Microsoft Excel®

and in process simulators. This work serves as guidance for engineering education researchers seeking a holistic, balanced approach to support students in developing process engineering deliverables while addressing ethical concerns associated with the integration of human-GenAI.

Keywords

Industry 5.0, Education 5.0, Generative Artificial Intelligence (GenAI), Chemical Process Design

Introduction

Generative AI (GenAI) has already been used in chemical process design within the context of engineering education [1], [2], with significant variability in implementation strategies and inconclusive, measurable outcomes regarding effectiveness. Thus, in some studies, although limited to small cohorts and pilot-scale initiatives, GenAI has been used to effectively support the development of various process engineering deliverables, including process flow diagrams, safety studies, and heat and mass balances [3]. Use strategies and targets also vary across students, with safety-related deliverables being generated more consistently when prompts are properly scaffolded [4]. Challenges persist, however, particularly in generating standardized process flow diagrams and control loops in piping and instrumentation diagrams. On the other hand, heat and material balances can be generated efficiently when prompted, but require significant refinement of calculations [5]. Despite these early initiatives documented in the literature, the research, to our knowledge, underscores the need to define effective prompt engineering [1] and comprehensive assessment strategies in chemical engineering education.

The effectiveness of GenAI is, therefore, task-dependent and depends on the level and complexity of scaffolding and explicit constraints defined in the learning/search process. GenAI tools consistently provide valuable content development, structured information generation, and guided feedback. However, these tools exhibit notable limitations when applied to tasks that require complex logical “thinking”, structuring and/or advanced technical problem-solving without adding external validation. This trend is evident across various engineering education contexts, including chemical engineering problem modelling, and design-oriented learning activities [6], [7], [8]. GenAI has also been studied as a supplementary aid for undergraduate design tasks, such as equipment sizing. These findings reinforce that GenAI's primary strength lies in assisting with step structuring, providing high-level alternatives, and organizing reports or assignments, rather than functioning as an autonomous designer [9]. Moreover, GenAI adoption is still at a general level, hence, details shall be provided regarding governance rules, scaffolding strategies, verification routines, and assessment instruments to properly align GenAI outcomes with the course learning outcomes, which remain not fully discussed in the current literature. For instance, research examining the acceptance and use of chatbots in engineering classrooms demonstrates significant variability in learner adoption and even trust, indicating that implementation effectiveness is influenced by both course design and expectations [3]. Assessment approaches also differ substantially across studies; while some implementations employ structured evaluation methods, many rely primarily on student perceptions (surveys) or participation indicators. This unstandardized reliance restricts comparability across educational contexts and weakens the validity of claims regarding achieved learning outcomes [8], [9].

The previous insights motivate the central argument of our study: GenAI integration in chemical process design should be approached as a framework and workflow problem. Chemical process design is particularly prone to errors that propagate across interdependent deliverables, such as process flow diagrams (PFDs)/piping and instrumentation diagrams (P&IDs), heat and material balances, and safety or economic analyses. This workflow requires mechanisms for verification/validation and alignment with established standards/codes. This study proposes a deliverable-anchored framework that operationalizes constrained GenAI use through prompt scaffolding, validation gates, and error-based feedback. The application of our framework is presented in the context of a hydrogen-from-biomass course project. This approach intends to address limitations, including uneven reporting of implementation mechanics, inconsistent validation techniques, and limited outcomes-based assessments that can be compared across cohorts and course designs. [3], [6], [7], [10].

In this work, we first conducted a review and bibliometric mapping of GenAI use in engineering/chemical engineering education to identify how current studies report application areas and evaluation approaches, and to identify where process-design-specific deliverables and validation practices are specified in literature. Second, we translated these insights into a deliverable-anchored framework and demonstrated its application through a course-based case study in a chemical engineering undergraduate course using a hydrogen-from-biomass design project. In the case study, the framework is operationalized through a validation-gate workflow, a deliverable-level prompt module that contrasts high-risk open-ended with scaffolded constrained prompting, and an error taxonomy that allows for consistent feedback and assessment across teams and deliverables.

Our contribution lies in identifying how GenAI is currently reported in engineering education and showing that process design deliverables (PFD/P&ID, balances, safety/control, sustainability, and economics) are often underspecified, making implementations difficult to replicate and generalize. Furthermore, we propose a pragmatic roadmap for chemical process design courses to link each process deliverable to scaffolded prompts, required validation evidence, and assessment criteria that reward traceability and responsible use, supporting Education 5.0/Industry 5.0 goals of human-centered and accountable decision-making.

Methods

Systematic review

Our systematic review reports describe the use of GenAI in undergraduate chemical/process engineering design education, aiming to capture the domains of implementation, their corresponding mechanisms, cross-validation methods, assessment and outcome metrics, and risks and/or ethical concerns. Our PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses)-ready Scopus Advanced Search strings were tuned to 2021-2026. The identification queries Q (with Advanced query option ON) were refined with ChatGPT 5.2 Thinking and include:

- Q1: High-recall core search (primary identification query)

```
TITLE-ABS-KEY(
("generative AI" OR genai OR "generative artificial intelligence"
 OR "large language model" OR "large language models"
 OR llm OR llms
 OR "foundation model" OR "foundation models"
 OR chatgpt
 OR "gpt-3" OR "gpt-3.5" OR "gpt-4"
 OR "AI chatbot" OR "AI chatbots"
 OR "language model" OR "language models")
AND
(educat* OR teach* OR learn* OR pedagog* OR curriculum OR course* OR classroom OR student* OR
capstone OR "design course" OR "project-based" OR "problem-based")
AND
(engineering OR "chemical engineering" OR "process engineering")
AND
(design OR "process design" OR "conceptual design" OR "process synthesis" OR flowsheet* OR
flowsheeting OR "process simulation" OR "process simulator" OR "process model" OR "process
models" OR "process diagram" OR "process diagrams")
)
```

This query seeks engineering education papers in chemical/process engineering that discuss the use of generative AI/LLMs (e.g., ChatGPT/GPT-4) in the context of design-oriented topics such as process design, flowsheets, process simulation, or process modeling/diagrams. It will capture conference and journal papers describing the use of ChatGPT/large language models (LLMs) in *any engineering design course* that mentions process design/simulation/flowsheets. It might miss papers that use a specific simulation tool or tool terms (e.g., Aspen HYSYS, Aspen Plus, PFD/P&ID) unless they also include one of the "process simulation / flowsheet / process diagram" terms. It may include some non-chemical engineering items, since bucket 3 includes plain "engineering" and bucket 4 includes plain "design."; the strict domain within chemical engineering was then defined with the SCOPUS tool.

- Q2: Chemical-process-design focused query

```
TITLE-ABS-KEY(
  (
    "generative AI" OR genai OR "generative artificial intelligence"
    OR chatgpt
    OR "large language model" OR "large language models"
    OR "language model" OR "language models"
    OR llm OR llms
    OR gpt OR "gpt-3" OR "gpt-3.5" OR "gpt-4"
    OR "AI chatbot" OR "AI chatbots" OR chatbot OR chatbots
  )
  AND
  (
    educat* OR teach* OR learn* OR course* OR curriculum OR capstone OR "design project"
  )
  AND
  (
    "chemical engineering" OR "chemical process" OR "process engineering"
  )
  AND
  (
    "process design"
    OR flowsheet*
    OR "process flow diagram" OR "process flow diagrams"
    OR pfd
    OR "piping and instrumentation diagram" OR "piping and instrumentation diagrams"
    OR "p-id" OR "p-ids"
    OR "process simulation"
    OR aspen OR hysys
    OR "heat and material balance" OR hmb
    OR "mass balance" OR "energy balance"
```

```
  )
)
```

This query captures papers where the title/abstract/keywords mention GenAI/LLMs (e.g., ChatGPT/GPT/LLMs) *and* an education context, chemical/process engineering,as well as terms related to process design coursework, such as PFD or P&ID, or simulation tools like Aspen HYSYS.

- Q3: Process diagrams plus GenAI query

```
TITLE-ABS-KEY(
  (
    "generative AI" OR genai OR "generative artificial intelligence"
    OR chatgpt
    OR "large language model" OR "large language models"
    OR "language model" OR "language models"
    OR llm OR llms
    OR gpt OR "gpt-3" OR "gpt-3.5" OR "gpt-4"
  )
  AND
  (
    pfd
    OR "process flow diagram" OR "process flow diagrams"
    OR "piping and instrumentation diagram" OR "piping and instrumentation diagrams"
    OR "p-id" OR "p-ids"
    OR flowsheet*
    OR "process diagram" OR "process diagrams"
    OR "control structure" OR "control structures"
    OR "process control"
  )
  AND
  (
    educat* OR student* OR course* OR classroom OR capstone OR "engineering education"
  )
)
```

This query finds papers describing the use of generative AI/LLMs (e.g., ChatGPT/GPT) in educational settings related to process engineering deliverables, such as PFDs, P&IDs, flowsheets, process diagrams, and control structures. It includes papers on chemical/process engineering education that use GenAI to generate/interpret flowsheets, PFDs, P&IDs, etc. It may miss papers that don't explicitly mention those diagram/control terms. It may include any discipline that uses those terms in an education context (unless a chemical/process engineering constraint is added on SCOPUS).

- Q4: Engineering education plus ChatGPT query

```
TITLE-ABS-KEY(
  (
    chatgpt
    OR gpt OR "gpt-3" OR "gpt-3.5" OR "gpt-4"
    OR "large language model" OR "large language models"
    OR "language model" OR "language models"
    OR llm OR llms
    OR "AI chatbot" OR "AI chatbots" OR chatbot OR chatbots
    OR "generative AI" OR genai OR "generative artificial intelligence"
  )
  AND
  (
    "engineering education"
    OR educat* OR teach* OR learn* OR pedagog*
    OR curriculum OR course* OR classroom
  )
)
```

This query finds papers that discuss generative AI/LLM tools (like ChatGPT/GPT/LLMs/chatbots) in an educational context, especially when the paper explicitly frames it as engineering education. Because the second bucket includes general education terms (educat*, course*, classroom, etc.), it can return engineering-education papers, *and also* general higher-education papers that mention "engineering" only lightly (or sometimes not at all), as long as they mention GenAI plus education in title/abstract/keywords. Chemical engineering articles have been constrained.

Inclusion criteria included an undergraduate engineering education context (journals and conference papers only), with emphasis on chemical/process engineering and design, explicit use of GenAI/LLMs, and a sufficient use description. Exclusion criteria included non-undergraduate contexts and domains not associated with chemical process design outcomes. Screening proceeded in two stages: (1) title/abstract/keyword screening; and (2) full-text screening, where studies that appeared to meet inclusion criteria or were ambiguous from abstract-only information. Inclusion and exclusion criteria were manually executed.

Abstract-based evidence was extracted from SCOPUS export files, including titles, abstracts, and keywords. Each study was manually clustered based on its design domain and deliverables, implementation mechanism, validation practices, assessment, and risks/ethics concerns. When only abstract-level information was available, records were assigned as reported/not reported/unclear based on the abstract. "Not reported" was not interpreted as the absence of practice, but as not captured by the abstract instead.

The software VOSviewer was used to cluster keyword and term co-occurrence relations from the retrieved **.ris* files to support domain-level mapping and to contextualize how GenAI applications and evaluation themes are distributed across the chemical engineering education literature.

We use bibliometric mapping and abstract-based evidence to derive design requirements for our framework; the framework is then contextualized in the course case study as a worked example and serves as the basis for future implementation.

Framework development

The review outputs allowed us to identify (i) the frequency with which GenAI use cases are reported in engineering education and process design in terms of deliverables and proper jargon, (ii) the implementation mechanics described across instructors and courses, and (iii) the validation practices included in the studies. These findings led us to design a framework that seeks deliverable-anchored integration and validation to make it suitable for standard chemical process design. The framework is then structured as a roadmap that combines (1) a prompt-engineering strategy, (2) a validation-gate workflow, and (3) an error taxonomy that supports consistent review, validation, and feedback. Figures 1 and 2 show the workflow gates and error taxonomy, respectively.

Course-based case study

The worked example described in this study is situated in a third-year undergraduate design course in our department of Chemical Engineering. This design course is delivered in the semester preceding the fourth-year capstone course. Students produce process deliverables culminating in a log report and updated process diagrams. Table 1 illustrates the course structure in terms of deliverables, as well as the expected technical and soft skills gained per deliverable.

Table 1. Course Structure and Gained Skills per Deliverable

Deliverable	Skills	
	Technical	**Soft**
Process drawings	Process description Process simulation/heat and material balance Process flow diagram Piping and instrumentation diagram	Teamwork and collaboration Problem-solving and critical thinking Technical writing Time management Decision-making/leadership Ethical judgment Presentations – NONE Communication with clients - NONE
Sizing and safety	Equipment sizing Hazard Identification (HAZID) study	
Circular economy	Capital Expenditures (CAPEX) and Operational Expenditures (OPEX) Life Cycle Assessment (LCA)	
Strategies report	-	

Teams were asked to design a renewable hydrogen production process from biomass. A representative team -for this study- developed a "Process Flow Diagram for Production of Renewable Hydrogen from Biomass," and scoped detailed instrumentation and control reasoning to the pyrolysis unit. Their control narrative explicitly prioritizes temperature, flow, and level loops and links level control to safety concerns.

The deliverables used as GenAI integration anchors include:

- Deliverable 3, which requires comparative LCA framing and discussion, circularity strategies for the selected process, and an economic analysis implemented in Excel/MATLAB/Python, with submission of calculation files and the report.

- Deliverable 4 (Log report) requires compilation of team strategies/workflow decisions and updated PFD/P&ID with reflections and recommendations.

The GenAI-supported design workflow with validation gates is shown in Figure 1.

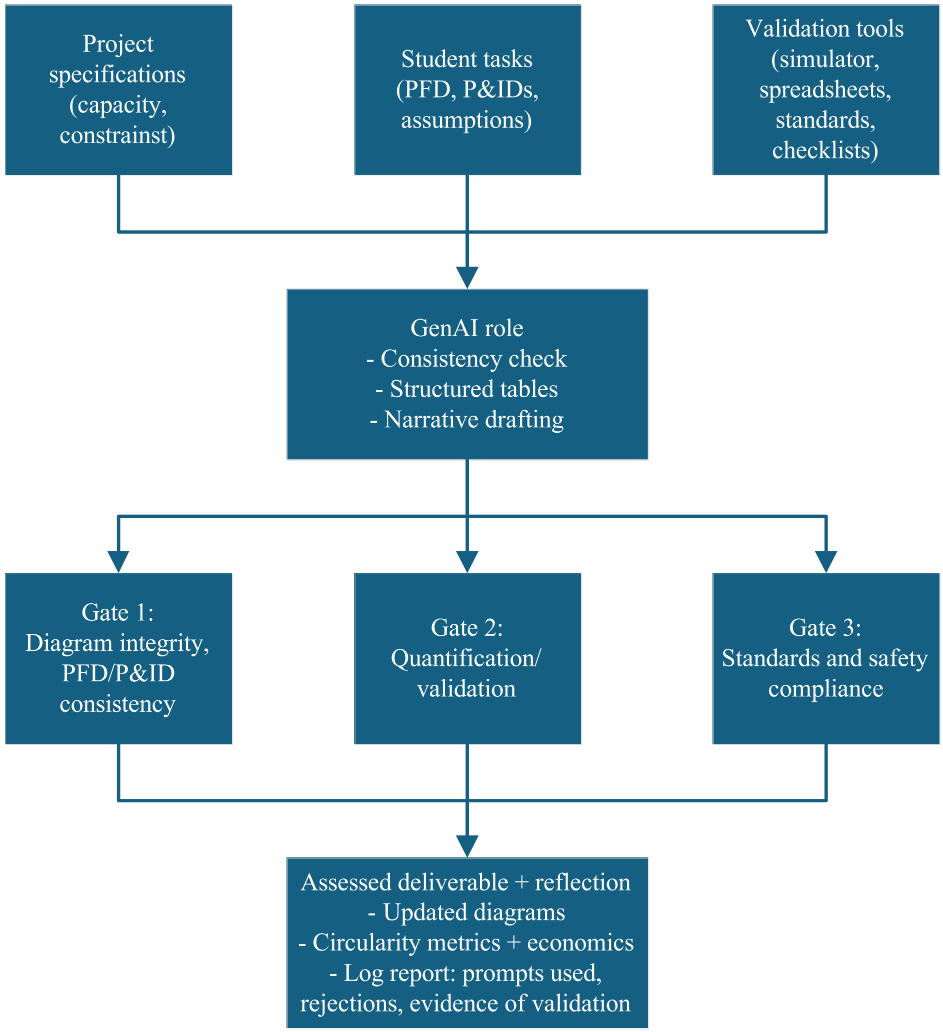


Figure 1. GenAI-supported workflow with validation gates

As can be seen in Figure 1, deliverables are evaluated against three validation gates: (Gate 1) representation integrity (PFD/P&ID consistency and completeness), (Gate 2) quantitative validation (agreement between a process model simulator and independently implemented calculations such as spreadsheets), and (Gate 3) standards and safety compliance (codes/standards alignment). Evidence for each gate is then recorded through submitted files (updated diagrams, calculations, and supporting checklists) and through team reflection logs included in each deliverable.

Figure 2 defines the error taxonomy used to review and classify GenAI-related issues into diagram/representation errors, quantitative/computational errors, safety/standards/compliance errors, and evidence/traceability errors. For each team and deliverable, reviewers (teaching team and subject expert matter) record (i) which validation gates are satisfied, (ii) instances of each error class, and (iii) whether the team's log report document mitigation actions (e.g., prompt constraints, cross-validations with references). These records maximize consistent comparison across teams and deliverable types and enable reporting of both performance and process measures.

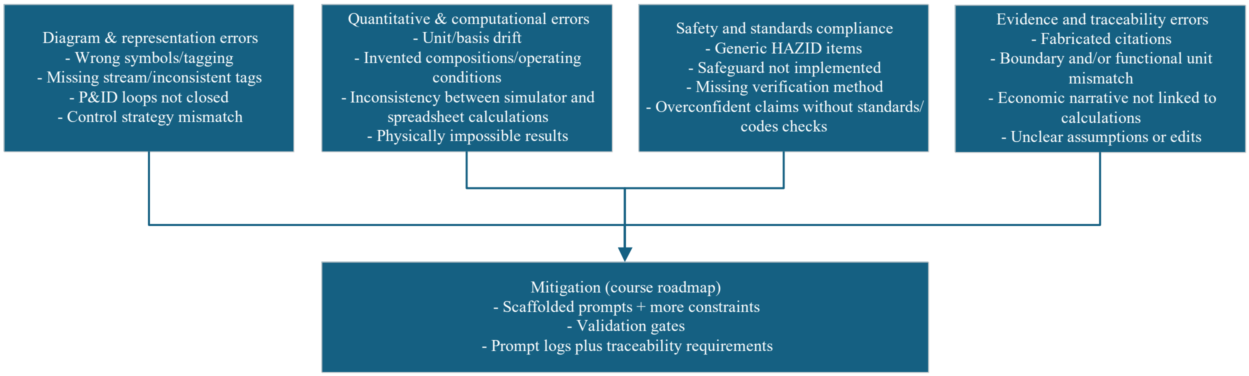


Figure 2. Error Taxonomy Used to Reviewing, Coding, and Classifying Gen-AI-related Issues

GenAI use in the course will be implemented through the workflow presented in Figure 1. Students will be encouraged to use GenAI primarily for troubleshooting and internal consistency checks/cross-validation, and, if required, drafting narratives based on provided/validated calculations and/or draft deliverables. Hence, prompts will be scaffolded by requiring teams to use inputs based on their own deliverables." Open-ended prompts that require full development of deliverables will be flagged as 'high risk' and discarded. The prompting sequence will consist of constraining the task and providing effective prompts as inputs, requesting structured outputs with clear fields, and performing an audit prompt when evidence is believed to be missing. Teams must document both accepted and rejected outputs, along with the validation evidence used (in the reflection section of each deliverable and in the final log report). In future offerings of our design course, the framework will be implemented longitudinally to evaluate product outcomes (deliverable quality and compliance) and process outcomes, aligned with the Education 5.0/Industry 5.0 emphasis on responsible human–AI teaming.

Our workflow combines circular-economy principles with design competencies and is supported by an assessment structure that enables GenAI validation and alignment with established standards. In the context of Education 5.0, the roadmap emphasizes team collaboration, enhances robustness (and trust) through scaffolded prompting, and strengthens reconciliation between the simulator and the spreadsheet, as well as evidence/validation logging across multiple validation stages. For Industry 5.0, the roadmap aligns deliverables with professional standards for safe and sustainable plant design by establishing safeguards, requiring justification and validation against standards and codes, and explicitly documenting provenance/assumptions for quantified claims regarding economic and sustainability outcomes.

Results

Systematic review

The clusters associated with Q1 (High-recall core search) for a total of 90 found documents are shown in Figure 3.

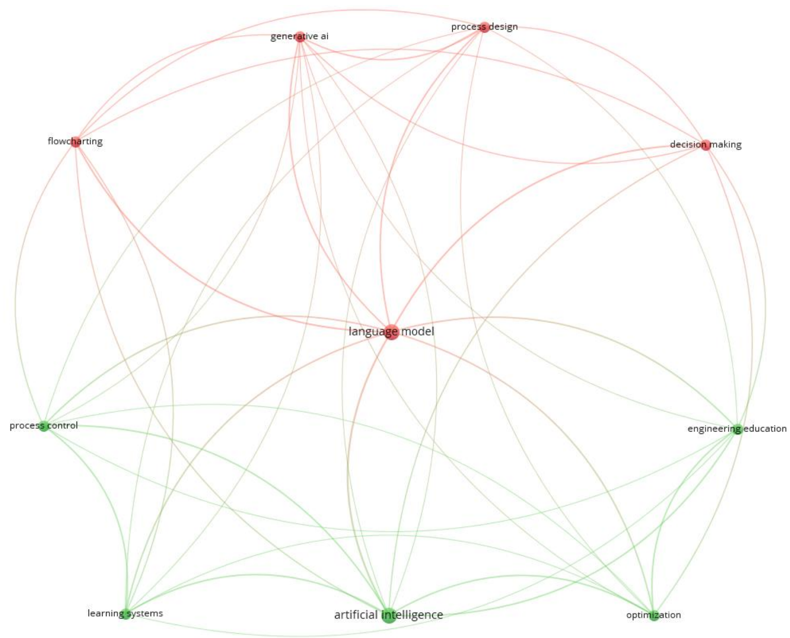
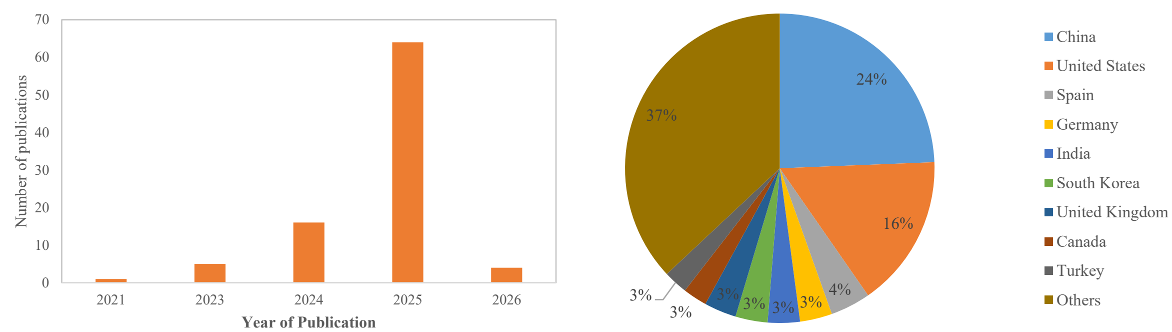


Figure 3. Bibliometric Clusters for Q1 (High-recall core search)

The corresponding number of studies per year and country distribution are shown in Figures 4a and b, respectively.



(a) (b)

Figure 4. Q1 - (a) Number of studies per year, (b) Number of studies per country

The number of studies found for Q2 (Chemical-process-design focused query) were 6 (3 in 2024 and 3 in 2025). Similarly, two documents were found for Q3 in the years 2025 and 2026 (Process diagrams plus GenAI query).

The clusters associated with Q4 (Engineering education plus ChatGPT query) for a total of 1,033 found documents are shown in Figure 5.

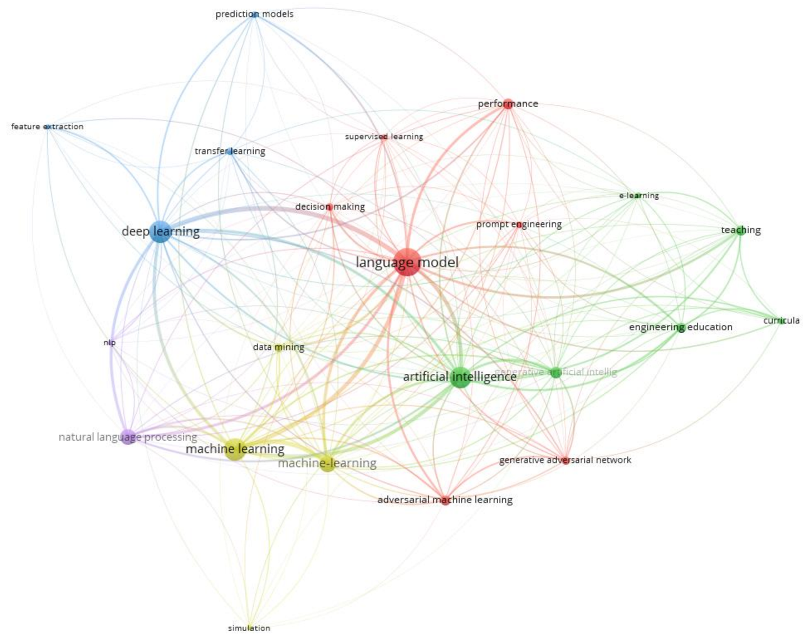
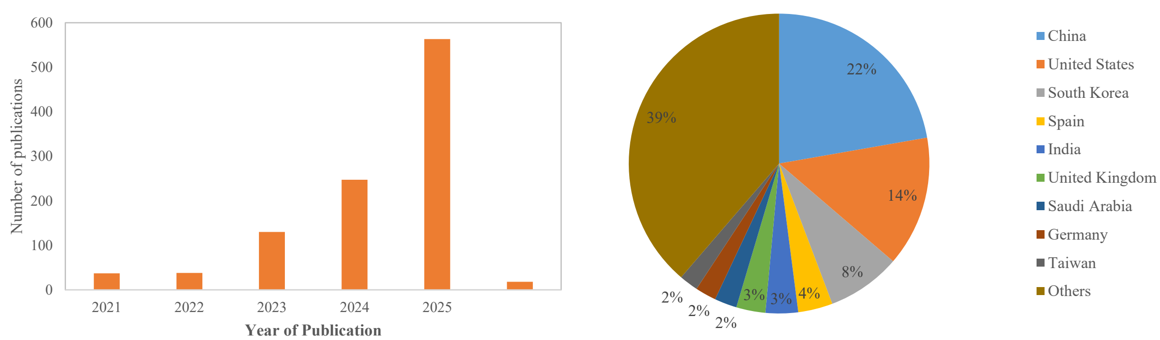


Figure 5. Bibliometric Clusters for Q4 (Engineering education plus ChatGPT query)

The corresponding number of studies per year and country distribution are shown in Figures 6a and b, respectively.



(a) (b)

Figure 6. Q4 - (a) Number of studies per year, (b) Number of studies per country

SCOPUS metadata was ‘cleaned’ by removing duplicates, enabling compilation of publications from 2021 to 2026. Explicit mention of process design deliverables was rare in the abstracts, with only 17 of 90 records (18.9%) directly referencing chemical, process engineering, or process design. Process design-specific jargon, such as flowsheets or PFD/P&ID, appeared in only 1 record (1.1%), as did process simulation tools, such as Aspen HYSYS. Safety- and control-related process design terms were similarly infrequent, each appearing in 2 records (2.2%). Implementation mechanisms were more commonly reported (29/90, 32.2%), as were educational uses such as coding or programming support (22/90, 24.4%), writing or feedback support (22/90, 24.4%), chatbot or tutor framing (19/90, 21.1%), and prompting or prompt engineering (19/90, 21.1%).

Course policy language, including guidelines and integrity rules, was also uncommon (10/90, 11.1%). Evidence related to validation was inconsistently reported at the abstract level, appearing in 31 records (34.4%). Assessment instruments were rarely mentioned, except for surveys, questionnaires, Likert-scale cues (12/90, 13.3%), validated-scale cues (13/90, 14.4%), and explicit rubric-based assessment (1/90, 1.1%). Risks and ethics cues, including bias, privacy, integrity, hallucinations, and safety concerns, were present in 32 records (35.6%).

Course-based case study

In practice, the most reliable learning outcomes were achieved when GenAI was used to structure well-defined work products such as tables or checklists, (ii) diagnose inconsistencies, and (iii) draft a narrative from verified values (provided detailed input from the user). On the other hand, the highest-risk prompts occurred when GenAI was prompted to "generate" complete engineering tasks without being anchored to student-generated evidence. Thus, sound deliverables can be produced when prompts are scaffolded by prior deliverables, and in other cases, when consistently troubleshooted.

To make this operational for students, we classified prompts by risk level: "high-risk prompts to avoid" (used as negative examples in instruction) and "scaffolded prompts to use" (paired with validation requirements). Some examples of different prompts per deliverable are shown as follows (created with ChatGPT, validated by the authors):

<u>Deliverable 1 – Process Flow Diagram and Piping & Instrumentation Diagram</u>

- Example 1: Bad prompt

  ```
  "Create a complete process flow diagram for a biomass-to-hydrogen plant with all unit operations, stream numbers, and operating conditions."
  ```

  Typical bad outcomes (Figure 2 error classes):

  - Evidence/traceability errors: invented temperatures/pressures/compositions with no source/reference
  - Representation errors: nonstandard equipment naming/tagging; missing key connections (e.g., purge handling, condensate separation)
  - Quantifiable errors: contradictory basis (dry vs wet biomass, mass vs molar) embedded in the narrative

  To avoid: unconstrained flowsheet generation.

- Example 2: Bad prompt

  ```
  "Add realistic values for stream compositions and flowrates throughout the plant."
  ```

  Typical bad outcomes:

  - plausible numbers that do not close; hidden assumptions inserted without labels ("assume 80% conversion…")

To avoid: not providing key constraints/assumptions.

- Example 3: Scaffolded prompts (good) tied to Gate 1 checks

```
"Audit our biomass→H₂ PFD for internal consistency using only the tag list and stream list pasted below. Do not invent tags, units, or values. Output an issue log: missing stream endpoints, duplicated tags, inconsistent stream naming, and missing utilities (steam, cooling water, power)."
"Using only the block description below (sections: feed prep → conversion → cleanup → shift → H₂ separation → compression), generate a checklist of required interfaces between sections (inputs/outputs/recycle/purge/waste). Mark any interface as 'UNKNOWN' if not stated. Do not propose new units."
"Rewrite our stream naming convention and equipment-tagging convention so it is consistent (e.g., S-###, V-###, HX-###). Use only the tags I provide. Provide a mapping table: old → new."
```

- Validation requirement (what students had to do to "count" it)
  - Gate 1 evidence: the issue log must be verified against the diagram, and fixes must appear in the revised PFD.
  - Any prompt output that introduced a new unit/stream/value was marked "rejected" unless backed by external reference or simulation.

<u>Deliverable 1 - Heat & Material Balance and simulation cross-validation</u>

- Example 1: Bad prompt

```
"Calculate the complete mass and energy balance for a biomass-to-hydrogen process."
```

Typical bad outcomes:

  - Quantification/computational errors: basis drift (kg dry biomass vs kg wet biomass), inconsistent units, energy terms mixed (LHV/HHV)
  - Evidence errors: invented syngas compositions and conversions without citing model/simulator/literature

- Example 2: Bad prompt

```
"My balance doesn't close—fix it and tell me what the right numbers should be."
```

Typical bad outcomes:

  - GenAI "fixes" by silently changing assumptions, which hides the real learning (debugging) and breaks traceability.

- Example 3: Scaffolded prompts (good) tied to Gate 2 checks

```
"Using only the simulator stream table pasted below (Stream ID, total flow, component flows, T, P), generate a unit-by-unit mass closure table. Output columns: Unit | in (kg/h) | out (kg/h) | closure error | likely causes | required checks. If a required variable is missing, write 'MISSING'."
```

- Example 4: Good prompt (basis drift detector):

```
"Scan this calculation excerpt and flag any basis drift: wet vs dry biomass, as-received
vs dry ash-free, mass vs molar, hourly vs annual. Output a list of flagged lines and what
basis each line is using."
```

- Example 5: Good prompt (energy accounting guardrails):

```
"Check whether the energy accounting mixes HHV/LHV or inconsistent reference states. Do
not compute values; only flag inconsistencies and what additional definitions are needed."
```

- Validation requirement (what students had to do)
    - Gate 2 evidence: show simulator ↔ spreadsheet agreement for at least one section (e.g., shift + PSA), and document any discrepancies plus how they were resolved (unit conversion, component basis, moisture handling).
    - Students were encouraged to include "debug trails" (what failed, what was corrected), matching your abstract's "deep troubleshooting required."

<u>Deliverable 2 – Safety and Control Loops</u>

- Example 1: Bad prompt

```
"Write a complete HAZOP for the entire biomass-to-hydrogen plant including safeguards."
```

Typical bad outcomes (Figure 2):

- Safety/standards errors: generic deviations not tied to a node; safeguards not implementable with listed equipment
- Traceability errors: no mapping to tags/streams, no verification method

- Example 2: Bad prompt

```
"Design the entire control system for the plant and specify control loops."
```

Typical bad outcomes:

- loops that sound sensible but do not close; missing final elements; inconsistent control objectives

- Example 3: Scaffolded prompts (good) tied to Gate 3 checks

```
"Using only this node description (tag, service, materials, T, P, upstream/downstream
streams) and the safeguards present in our P&ID list, generate a HAZID table: Deviation
| Causes | Consequences | Existing safeguards (tag) | Recommendation | Verification
method. Do not invent safeguards. If a safeguard requires new hardware, label it 'design
change.'"
```

- Example 4: Good prompt

```
"For each recommendation below, state whether it is implementable with our current
equipment/instrument list. Output: recommendation | existing tag to attach | requires new
equipment? | requires new control logic? | verification method."
```

- Example 5" Good prompt

```
"Given the list of proposed loops (CV, MV, sensor tag, final element tag), check whether each loop has sensor/controller/final element and whether MV is physically plausible (valve speed, heater duty, compressor speed). Mark incomplete loops and list what is missing."
```

- Validation requirement
  - Gate 3 evidence: every safeguard must map to a diagram element or be explicitly labeled a design change; every recommendation must include a verification method (standard, calculation, or test).
  - This is how you make "safety fairly described" without letting GenAI invent systems.

Deliverable 3 - Circularity and Economics

- Example 1: Bad prompt

```
"Compare the LCA of biomass-to-hydrogen with SMR and provide GWP numbers with citations."
```

  Typical bad outcomes:
  - fabricated/weak citations, boundary mismatch, unjustified numbers
- Example 2: Bad prompt

```
"Estimate CAPEX, OPEX, and hydrogen selling price for our plant."
```

  Typical bad outcomes:
  - invented scaling exponents, inconsistent year dollars, totals that cannot be linked to student files
- Example 3: Scaffolded prompts (good) tied to traceability

```
"Using only our functional unit and system boundary statement, produce a comparison template that lists included/excluded stages and required inventory data. No numbers, no citations."
```

- Example 4: Good prompt (evidence-linked economics narrative)

```
"Write a results paragraph using only these spreadsheet totals pasted below. After each number, include the exact cell reference I provide. If any claim cannot be supported by pasted values, write 'NOT SUPPORTED.'"
```

- Example 5: Good prompt (assumption register generator):

```
"From the pasted assumptions list, generate an assumption register table: assumption | where used (deliverable) | impact direction | how to validate (simulator/standard/reference). Do not add assumptions."
```

- Validation requirement
  - Every numerical statement must be traceable to a spreadsheet/simulator output (or clearly labeled as an assumption with a validation plan).

- Comparative sustainability claims require boundary/functional unit alignment before any numbers are discussed.

Discussion

Systematic review

The search strategy for SCOPUS included a high-recall query (Q1) alongside a set of precision queries (Q2–Q2). Q4 (“Engineering education plus ChatGPT”) yielded 1,033 records, capturing the broader implementation of GenAI in engineering education. Q1 (“High-recall core search”) identified a design-focused corpus (n = 90), aligned with engineering design and process design terminology. The low counts for the specific queries—Q2 (n = 6) and Q3 (n = 2)—suggest that few studies explicitly reference GenAI use in undergraduate chemical process design, as defined by standard chemical process-engineering terminology. Additionally, assessment instruments are rarely mentioned in abstract metadata, and explicit rubric-based task assessments are inconsistently reported across studies. The lack of validation details in abstracts further highlights the need for a course-based roadmap that defines verification gates and assessment instruments in a deliverable-anchored format. Furthermore, risks and ethics cues appear in approximately one-third of engineering-education-focused abstracts, typically discussed in terms of general policies or responsible use, with almost no mention of mitigation measures. The rare mention of explicit safety or validation language in abstracts reinforces the need for structured human-in-the-loop verification workflows and clear traceability requirements when integrating GenAI into process-design deliverables.

Course-based case study

Our findings suggest that GenAI in process design courses should be used as an instructional design and assessment tool, rather than as a deliverable-producing tool, because process design deliverables are interdependent and errors might propagate from early to more consolidated versions/cross-validated deliverables. Our validation-gate approach offers a pragmatic way to teach GenAI use as an engineering practice, while supporting outcomes with evidence and validation (effective prompt engineering approach[1]). The major implications of our findings can be summarized as follows: (i) deliverable-specific scaffolding is crucial, (ii) GenAI should be primarily a checker/structuring assistant rather than a full deliverable generator (for which learning value is preserved), and (iii) assessment should reward verification and validation, rather than polished prompts (this supports the human-centred accountability emphasis of Industry 5.0 while ethical judgement is assessed). Appendix 1 includes the validation gates and error taxonomy by deliverable in the course roadmap.

Limitations and future work

Our review is constrained by the level of detail available in the abstracts. As a result, the abstract level is less reliable for reconstructing verification/validation workflows and assessment instruments. Hence, this approach may undercount practices that were implemented but not described in abstract metadata.

A second limitation is jargon inconsistency across the literature. Process design deliverables (e.g., flowsheets, PFD/P&ID, simulation platforms, safety analyses) are inconsistently named in abstracts, even when the underlying study is situated in design contexts. Consequently, process task retrieval and coding likely underestimate the true prevalence of GenAI use in process-design deliverables.

Reliance on a single primary database (SCOPUS) may omit relevant studies indexed elsewhere (e.g., IEEE Xplore, ACM DL, arXiv/SSRN) that are not fully captured by SCOPUS. Therefore, the findings should be interpreted as a systematic mapping of what is reported in the Scopus-indexed record during the study window only. Also, we expect future work to compare results generated by other chatbots, for generalization purposes.

In the future, we expect to refine our prompt library into a generalizable instructional resource (prompt templates and checklists) and investigate how GenAI integration can be aligned with program-level outcomes associated with Education 5.0 and Industry 5.0).

Conclusion

In this study we conducted a review to identify how GenAI is reported in engineering education and where process-design-specific deliverables, validation practices, and assessment instruments remain underspecified. These insights serve as a basis to develop a deliverable-anchored framework for GenAI integration in undergraduate chemical process design, operationalized through scaffolded prompting, three validation gates (diagram integrity, quantitative and validation, and standards/safety compliance), and an error taxonomy for consistent review. We illustrated the implementation of our framework with a hydrogen-from-biomass worked example, which could be incorporated into our design-based undergraduate third-year course, Team Strategies for Engineering designs. The corresponding results show that GenAI is most reliable when constrained to perform tasks such as checking, structuring, and drafting from detailed/verified inputs, and, consequently, that the highest risk occurs when prompted to generate complete deliverables without traceable evidence. The framework operationalizes the Education 5.0 and Industry 5.0 elements by emphasizing team-based workflows, transparency, alignment with standards, and accountable decision-making. Future course offerings will implement the framework longitudinally, requiring prompt logs and validation evidence to enable evaluation of deliverable quality and prompt verification/validation.

Acknowledgement

The authors would like to acknowledge the use of ChatGPT 5.2 Thinking (OpenAI) for supporting the structuring of this paper's content. The conceptualization, methodology, analysis, validation, and writing have been performed by the authors.

References

[1] S. Chakraborty and D. Galatro, “Assessing the Accuracy of ChatGPT in Describing Chemical Processes using Effective Prompt Engineering,” in *Canadian Conference of Engineering Education*, Kelowna, BC, Canada, 2023.

[2] D. Galatro and S. Chakraborty, “Chemical Process Design to meet Industry 5.0 competencies,” in *2025 ASEE Annual Conference & Exposition Proceedings*, ASEE Conferences. doi: 10.18260/1-2--56087.

[3] C. Honig, S. Rios, and A. Desu, “Generative AI in engineering education: understanding acceptance and use of new GPT teaching tools within a UTAUT framework,” *Australasian Journal of Engineering Education*, vol. 30, no. 1, pp. 80–92, Jan. 2025, doi: 10.1080/22054952.2025.2467500.

[4] S. Tassoti, “Assessment of Students Use of Generative Artificial Intelligence: Prompting Strategies and Prompt Engineering in Chemistry Education,” *J. Chem. Educ.*, vol. 101, no. 6, pp. 2475–2482, Jun. 2024, doi: 10.1021/acs.jchemed.4c00212.
[5] J. Cordero, J. Torres-Zambrano, and A. Cordero-Castillo, “Integration of Generative Artificial Intelligence in Higher Education: Best Practices,” *Educ. Sci. (Basel).*, vol. 15, no. 1, p. 32, Dec. 2024, doi: 10.3390/educsci15010032.
[6] M.-L. Tsai, C. W. Ong, and C.-L. Chen, “Exploring the use of large language models (LLMs) in chemical engineering education: Building core course problem models with Chat-GPT,” *Education for Chemical Engineers*, vol. 44, pp. 71–95, Jul. 2023, doi: 10.1016/j.ece.2023.05.001.
[7] F. Caccavale, C. L. Gargalo, K. V. Gernaey, and U. Krühne, “Towards Education 4.0: The role of Large Language Models as virtual tutors in chemical engineering,” *Education for Chemical Engineers*, vol. 49, pp. 1–11, Oct. 2024, doi: 10.1016/j.ece.2024.07.002.
[8] Z. Zhang and Y. Chang, “LEVERAGING GENERATIVE AI TO ENHANCE ENGINEERING EDUCATION AT BOTH LOW-LEVEL AND HIGH-LEVEL STUDY,” in *2024 ASEE Mid-Atlantic Section Conference Proceedings*, ASEE Conferences. doi: 10.18260/1-2--49450.
[9] N. I. Schnatz and A. Bieri, “REFINING AN AI-INTEGRATED SCIENTIFIC WRITING COURSE FOR ENGINEERS: INSIGHTS FROM A DESIGN-BASED RESEARCH APPROACH,” Jun. 2025, pp. 2394–2406. doi: 10.21125/edulearn.2025.0663.
[10] Z. Y. Kong, V. S. K. Adi, J. G. Segovia-Hernández, and J. Sunarso, “Complementary role of large language models in educating undergraduate design of distillation column: Methodology development,” *Digital Chemical Engineering*, vol. 9, p. 100126, Dec. 2023, doi: 10.1016/j.dche.2023.100126.

Appendix 1 - Validation gates and error taxonomy by deliverable in the course roadmap

Deliverable (roadmap step)	**Primary learning focus**	**Primary validation gate**	**Secondary validation gate**	**Most common error classes**
D1 – PFD and early P&ID	Establish flowsheet logic, tags, streams, interfaces, initial control strategy	Gate 1: Diagram integrity (PFD/P&ID consistency, tagging, loop completeness)	Gate 2, only at template level if numbers are yet not final	Diagram/representation: nonstandard symbols or tagging, missing or inconsistent streams, incomplete loops Evidence/traceability: phantom units and/or values
D1 – Heat and Material Balance / Simulation	Quantitative reconciliation and basis management	Gate 2: Quantification / validation (simulator and spreadsheet agreement, closure checks)	Gate 1 (stream table must align with diagram) Gate 3 (only if safety limits referenced)	Quantitative/computational: basis drift, unit errors, physically impossible results Evidence/traceability: missing exports, silent assumption changes
D2 – Control and safety	Hazard identification to specific nodes; implementable safeguards	Gate 3: Standards and safety compliance (implementable safeguards; codes/standards check)	Gate 1 (safeguards must map to tags and loops) Gate 2 (safeguards sometimes require sizing and/or calculations)	Safety/standards/compliance: generic hazards, non-implementable safeguards, missing verification method/reference Diagram/representation: control not reflected Evidence/traceability: recommendation not tied to tags
D3 – Circularity/LCA, economics	Scope discipline, traceable economics, sustainability claims	Gate 2: Quantification / validation (economics, results traceable to spreadsheets) and Gate 3: scope boundary integrity for sustainability claims	Gate 1 (diagrams consistent with inventory boundaries where relevant)	LCA/circularity claims stated with functional units and boundaries, data gaps identified before numerical comparison. Economics narrative traceable to spreadsheet