Nomination for the **CTL Curriculum Innovation Award 2022**

Nominees: **The Graduate Certificate in Data Science for the Chemical Industry Team**:

Drs. Fani Boukouvala, Martha Grover, Andrew "AJ" Medford, J. Carson Meredith, and David Sholl

Table of Contents

February 6, 2021

CTL Curriculum Innovation   Award

Dear Award Selection Committee:

It is my pleasure to nominate a team of faculty in the School of Chemical & Biomolecular Engineering for recognition with the CTL Curriculum Innovation Award . Specifically, this team includes Professors Fani Boukouvala, Martha Grover, Andrew (AJ) Medford, J. Carson Meredith, and David Sholl. This nomination details their development of the   **Graduate Certificate in Data Science for the Chemical Industry** , which is Georgia Tech's first fully online graduate certificate.

The goal of this program is to bridge the gap between chemical engineering and computer science, providing chemical engineers with the basic programming and mathematical skill needed to grasp key concepts in data science and machine learning. It differs from existing online programs, such as the OMS in Computer Science or the OMS in Data Analytics in that teaches these techniques in the context of chemical engineering problems, rat her than abstract computing challenges or business case studies.

 This program is innovative in several ways. In addition to being GT's first fully online graduate certificate, it is the first program of its kind designed for professionals in the chemica   l industry. Indeed, companies (e.g,. 3M, Dow Chemical, and others) have partnered with GT to enroll their employees in these courses to provide skills in the emerging field of data science and data analytics. At the same time, the courses are also availale to our resident GT graduate students, as well as senior graduate students, who take the courses for GT credit towards their degrees as fully online courses, alongside the professional education cohort.

The certificate is built upon two core courses, C hBE 4745/6745 Data Analytics for Chemical Engineers, and ChBE4746/6746 Datadriven Process Systems Engineering, coupled with two elective courses from a list of options offered in -person at GT or online. The two core course sequence employs vertically-in tegrated teams of undergraduate, graduate, and industry students. This structure provides undergraduate students a chance to gain first -hand insight into graduate school and industry careers and provides industry students with the opportunity to mentor und ergrads and learn from grad students. Reviews from students are very strong, as outlined in this nomination package.   As Zak Kuiper of the Mosaic Company points out, "….. this program is first to the table in answering a need which most of the chemical industry is just now realizing they have. It has allowed me to develop as a professional significantly quicker than I

would have otherwise and has shaped how our organization has focused on the area of Data Science."

This curricular innovation is well-aligned with many strategic Georgia Tech and ChBE@GT directions. It provides virtual classroom connections between students at all levels (undergraduate – graduate – professional). It establishes Georgia Tech's leadership in data analytics aspects of the chemical process and product industries. It enhances our connections and relationships with industry, and provides a platform for scalable, global educational outreach.

In my twenty-one years on the faculty, we have often aspired to be national and global thought leaders, being too often fast followers. Nonetheless, during this period we have grown as an institution and are increasingly now driving innovation in key education and research sectors. In 2022, we are now global leaders in online teaching and learning of computer science. And this new program in ChBE@GT leverages this foundation created by our colleagues in the College of Computing and GT Professional Education to launch the first program of its kinds targeting data science in the chemical sector. The **Graduate Certificate in Data Science for the Chemical Industry** (DSCI) is not only a fabulous educational innovation, it is also a crucial thought - leadership platform, one that where we are far ahea d of the curve, offering programmatic innovations that other institutions are just starting to imagine. I am tremendously proud of our DSCI team– we are fortunate to have them as members of our community. I nominate them for this recognition with great enthusiasm and admiration.

Sincerely,

Christopher W. Jones, Ph.D.
John F. Brock III School Chair and Professor
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

I. **Description of the innovation, including the problem or student learning issue it addresses, the objectives of the innovation, the learning outcomes for the intended audience, and the approach taken.**

The **Graduate Certificate in Data Science for the Chemical Industry (DSCI) is Georgia Tech's first fully online graduate certificate**. By benefiting both off-campus and on-campus students, it creates a template for much broader innovation in graduate education at Georgia Tech that is likely to appeal to a very large group of students for whom a complete online MS is too large of a commitment. The DSCI program was designed in collaboration with partners in the chemical industry, who now fund the tuition costs of their engineers and scientists who are students in the program.

This certificate bridges the gap between chemical engineering and computer science, providing chemical engineers with the basic programming and mathematical skill needed to grasp key concepts in data science and machine learning. The curriculum builds on these foundational skills through examples specifically related to the optimization of chemical processes, providing chemical engineers with practical knowledge of how data science techniques can be applied to chemical processes.

The **team of nominees** is as follows:

- **Dr. Fani Boukouvala**, Assistant Professor, School of Chemical & Biomolecular Engineering
- **Dr. Martha Grover**, Professor and Associate Chair for Graduate Studies, School of Chemical & Biomolecular Engineering
- **Dr. Andrew (AJ) Medford**, Assistant Professor, School of Chemical & Biomolecular Engineering
- **Dr. J. Carson Meredith**, Professor and James Harris Faculty Fellow in the School of Chemical & Biomolecular Engineering and Executive Director of RBI
- **Dr. David Sholl**, Professor, School of Chemical & Biomolecular Engineering

A. Motivation

In the past decade, society has been transformed by a revolution in the collection and use of data. Over roughly the same period, the chemical process industry (CPI) has gone through a manufacturing renaissance in the US, with hundreds of billion dollars of construction on the Gulf Coast alone. The CPI employs more than 800,000 people in the US and created products worth more than $700 billion in 2017. Despite the vast scale of the CPI, which encompasses oil and gas, chemicals, consumer products, and pharmaceuticals, this sector badly lags behind other areas in taking advantage of data science. The CPI is an interdisciplinary field that employs chemical engineers, chemists, material science engineers and bioengineers, who are all needed to design, operate and optimize materials, formulations, processes and products. A 2018 McKinsey study estimated that incorporating Artificial Intelligence in operations within the global CPI

would realize $800 billion in annual value. However, according to industry feedback, simply embedding data scientists without CPI training in risky chemical manufacturing environments does not lead to meaningful change in operations, and oftentimes leads to distrust of data-science from domain experts. At the same time, the CPI's current employees have domain expertise and first-hand experience of the effects of digitalization, but lack the necessary data science background. ***A key need in the US CPI is development of a highly skilled workforce that combines domain expertise in industrial-scale chemicals processing with robust and adaptable data science skills.***

Despite the big-data revolution, the chemicals, materials, and products manufacturing industries have been more reluctant towards embracing these new technologies for a variety of reasons. The reason up to a few years ago was the lack of digitization, which is now a reality that has led to an abundance of operational time series data, text data, and even image and video data. However, unlike the advertising and social media industries, a misclassification or a false positive prediction when controlling the production of a chemical can have significant safety, health, and environmental consequences. As a result, these industries are in great need of individuals who have domain expertise coupled with a strong background in Data Science. This can be achieved both by hiring new employees with this interdisciplinary background, but also by retraining their existing domain experts with this new set of skills. **The program designed here by the School of Chemical & Biomolecular Engineering at Georgia Tech aims to fill this gap of (re)training individuals that form the core of the technical and R&D workforce in a broad swath of companies focused on commodity chemicals, specialty chemicals, consumer products, and many more.**

The need for this new certificate program was determined through discussions with ChBE industry partners and the external advisory board. Detailed discussions with industry partners including 3M and Dow indicated companies in the CPI are <u>not interested in sending their employees to a general data science program</u>, for a number of reasons:

- First of all, their employees are unlikely to have the *background* in mathematics and computing required for success in the program and for mastery of the material.
- More significantly, a generalized program would not provide the context needed to connect the chemical process industries with general data science approaches, limiting the potential positive impact of the program on the industry. This latter point has been made in a publication of the American Institute of Chemical Engineers as well [1].

Rather, our industry partners request a specialized program designed for the chemical process industries, and based on case studies from the chemical process industries. These internal conversations are consistent with published reports: "…it is much easier to train chemical engineers on data analytics topics rather than to train data scientists on chemical engineering topics" [2].

Nationwide, there is also no similar certificate program in data science. Northwestern University has a six-month data-science bootcamp, similar to Georgia Tech's professional education program. The University of Washington has the DIRECT program, which is focused on internal graduate students with content at the intersection of molecular phenomena and data science.

However, it is not open to external participants. The University of Buffalo has a MS program (full time/part time) for data science, but not with the specific focus on chemical engineering

Therefore, the innovation represented in the DCSI program is not only desired by industry, but is the first of its kind, offering a unique opportunity to employees and students alike.

B. Objectives

The objective of the DSCI program is to provide engineers in the chemical industry with knowledge of basic data science and machine learning techniques, and experience applying these techniques to specific problems in chemical engineering. This objective supports Georgia Tech's mission of using technology to advance the human condition.

This program is designed to benefit not only external participants working in the chemical industry, it is also open to existing Georgia Tech graduate students. Moreover, the two core courses in the program are open to GT undergraduates as well. This program brings the internal and external groups together, so that they can learn and benefit from each other. The younger generation students are expected to have theoretical domain expertise and better digital skills, but they can have a transformational experience by interacting with people with workplace skills (e.g, critical thinking, communication skills), practical domain knowledge, hands-on experience with real industrial data, and business experience in areas such as decision-making, and project and operations management.

The set of skills provided via this program are foundational requirements for employees to become competitive and impactful in the era of "Industry 4.0" and "Smart Manufacturing". This skill set is currently missing in CPI, and it is the main reason for the lag in adoption of state-of-the-art ML and AI tools that will modernize the manufacturing industry and make it competitive in the global markets. This program plays a key role towards the initiatives for reskilling and diversifying the US STEM workforce, which is a critical issue identified by NSF and the Business Higher Education Forum. In addition, the program aims to benefit society by educating students on how to use data-science ethically and responsibly to design and operate manufacturing and processing plants safely and sustainably, by maximizing energy efficiency and minimizing harmful waste.

This data certificate program supports the Institute's Strategic Plan in many ways. It establishes Georgia Tech's leadership in the data aspects of the chemical process industries, with the first data science certificate designed for the chemical process industries. The connections with industry facilitated by the certificate program will help to sustain and enhance ChBE's excellence in scholarship and research, through continuous injection of industry perspective and case studies, as well as facilitate future industrial research funding. The internal participants (GT students) benefit from their interactions with the external participants, fostering their broader perspectives on topics ranging from innovation to entrepreneurship to global awareness, and providing internship opportunities. The educational resources at Georgia Tech are now available to external participants worldwide, expanding access to high-quality resources at Georgia Tech.

C. Learning Outcomes

The curriculum is designed based on the following objectives:

<u>Objective 1:</u> Competence in Fundamentals of Data Science in Chemical Manufacturing
Students will develop and demonstrate competence in fundamental knowledge of data science relevant to chemical manufacturing.

<u>Objective 2:</u> Competence in Software Applications of Data Science in Chemical Manufacturing
Students will develop and demonstrate competence in using programming and software that uses data science to solve real-world problems in chemical manufacturing.

Graduates of the data science certificate will be able to

- understand how data science is relevant to chemical manufacturing.
- use data science programming and software to solve real-world problems in chemical manufacturing.
- lead data-driven efforts in their companies.

D. Approach

The Graduate Certificate in Data Science for the Chemical Industries is designed to provide technical depth and industry context in a format and timing that is accessible to professionals working full time. Designed to be completed in one to two years, the DSCI program consists of six hours of core coursework on foundational data science methods, with a strong emphasis on applications in the chemical process industry. An additional six hours of electives provides students with the opportunity to focus on their specific areas of interest and many are selected from within Georgia Tech's highly successful online master's degree in analytics.

Courses coincide with the Georgia Tech Academic Calendar. Core courses are offered in Georgia Tech's fall and spring semesters, while some elective courses are also available in Georgia Tech's summer semester. Students can complete course assignments during the hours that work best for them, but they must finish each course within the required timeframe. Each course has built-in deadlines and assessments along the way to make sure students stay on track. Each course is just as rigorous as its on-campus equivalent. Students in Georgia Tech's online programs who work full-time are typically able to complete one course per semester.

Both core courses are offered online in a semester format, making them available simultaneously to on-campus and off-campus students. Faculty teaching each course hold regularly scheduled virtual office hours and are also available to students asynchronously. Other best practices for ensuring student access to faculty have been adopted upon advice from successful existing programs in GTPE. For example, faculty encourage student teams to invite them to their regular group meetings, where they discuss progress and feedback on their semester projects. Students are also given the option to include professors in their Teams sites, where faculty can easily answer questions and provide feedback.

Internal graduate students take on-campus electives in addition to the online core courses. The online content for the two core courses was developed based on video recording of lectures with the support of the Office of Distance Learning and Professional Education. The two core ChBE classes are offered online, and a subset of the electives are also offered online. Students taking the online courses are able to view videos of the lectures and discussions, and to participate in group project efforts through remote communication with group members. On-line students have access to campus learning tools including Canvas and related applications, such as Piazza and video conferencing software, which facilitate group interactions.

D-1: Core Courses

1. **Data Analytics for Chemical Engineers (ChBE 4745/6745)**

This course establishes a foundation for handling data and understanding the difficulties that arise in trying to curate and learn from manufacturing systems data. It introduces different classes of data that are commonly encountered such as time series and property data. It introduces different classes of learning problems and the techniques for learning from the data. Students learn basic programming skills in Python and are introduced to key libraries for machine learning and optimization. It provides a map for students to connect the problem they are trying to address to the class of learning techniques that could be appropriate to use.

2. **Data-driven Process Systems Engineering (ChBE 4746/6746)**

In chemical process engineering and manufacturing, decision-making requires the formulation and solution of complex optimization problems. This course first introduces students to the basics of optimization, such as formulations, linear, nonlinear and mixed-integer programming. The second and main part of the course focuses on case-studies where the optimization needs to rely on data (from experiments, historical databases or simulations). In this part, the course introduces key concepts with respect to sampling and design of experiments, data-driven modeling and regression, model validation and optimization using data-driven models. Data-driven decision making is discussed in the context of applications from the chemical manufacturing and processing industry, such as: process operation and control, process design, process and material design, multiscale process modeling, monitoring, planning, scheduling and supply chain optimization. Overall, by the end of the course the students will have a good understanding of how to couple data-analysis with optimization for decision-making in the chemical process industries.

D-2: Innovations in Core Courses: ChBE 6745 and 6746

Semester-long Project

A key component of both ChBE 4745/6745 and ChBE 4746/6746 is the semester-long project. The project is completed by "vertically integrated" groups of 4-5 students (typically 1 from industry, 2 grad, and 2 undergrad). The project starts very early in the semester when the industry and graduate students are required to provide the dataset for the project and define the goals. Instructors provide feedback on the project definitions and allow teams to revise the

dataset and goals based on the feedback. Teams are strongly encouraged to work on real datasets and problems from industry or graduate research.

This project-driven course design and its focus on real problems helps motivate students and provides concrete examples of how the techniques they learn can be applied to real problems.

In ChBE 4745/6745 the project is broken into multiple phases (data preparation, baseline model development, model improvement, and final report/presentation). Similarly, in 4746/6746 the phases include: data preparation or collection, Optimization Formulation, Development of Optimization Method(s)/Algorithm(s), and final report/presentation. In both courses, teams receive feedback from instructors at regular checkpoints. Students can retroactively revise any portion of the project to yield the highest quality results by the end of the semester. An example data set for the project is shown below.
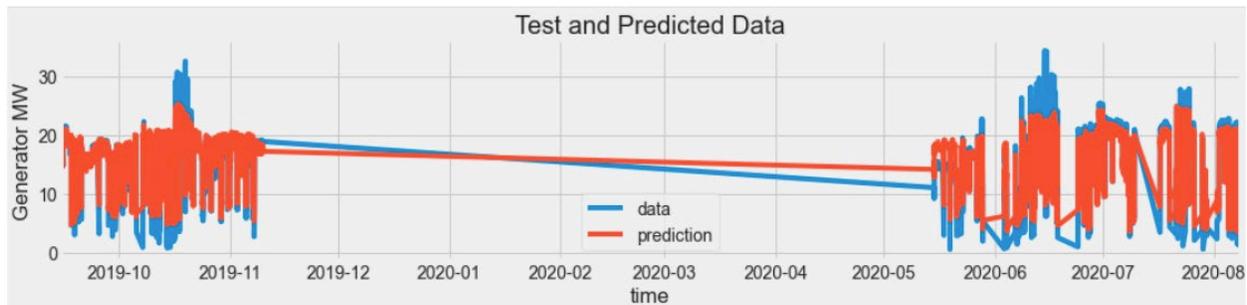

Fig. 1: Actual (blue) and predicted (red) data for power generation by a turbine generator at The Mosaic Company. Predictions are based on the operating parameters of the generator and an XGBoost regression model.

Project topics range widely but must be related to chemical engineering or chemistry, and often come from real industry or research projects. In ChBE 4745/6745, some examples include the prediction of polymer solubility, prediction of battery lifetime from early cycle data, analysis of mass spectroscopy data for nanoparticle size detection, prediction of the outputs of unit operations in a cumene production process, and prediction of power generation by a turbine generator. In ChBE 4746/6746, some of the topics studied include: optimization of organic electronics, the development of an optimal "Isotherm Modeler" tool, calibration of plasma mass spectrometry data, optimization of hippocampal neurons, optimization of vanadium flow battery design, and optimization of phosphoric acid production, to name a few.

 Ideally, projects provide an opportunity for industry students to bring in a real problem and gain tangible value from the project. A specific example is the prediction of turbine power generation project from Fall 2020, led by Zak Kuiper from The Mosaic Company, which produces phosphate fertilizers. In this project, students utilized a variety of regression models to predict the power output of a turbine generator as a function of operating parameters. **Zak was able to provide real process data, allowing other students in the group to gain experience working with a "data historian" and overcoming challenges associated with cleaning and**

processing real industry data. The results, shown in Fig. 2, showed strong predictive capability, and motivated Zak to continue working on this project in 4746/6746, and **ultimately using the results as the foundation of an internal project at Mosaic that has an estimated gross return of 8 million dollars**.

The same student was also able to continue working on a similar topic in ChBE 6746, where he and his team studied the optimization of the supply chain of phosphate production. In their model
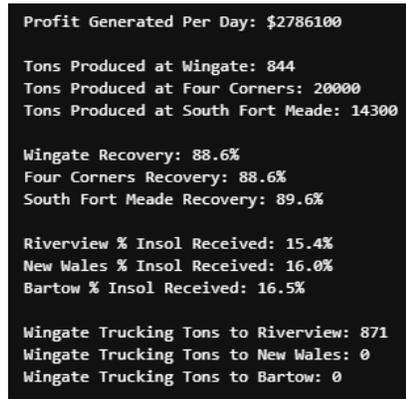
```
Profit Generated Per Day: $2786100

Tons Produced at Wingate: 844
Tons Produced at Four Corners: 20000
Tons Produced at South Fort Meade: 14300

Wingate Recovery: 88.6%
Four Corners Recovery: 88.6%
South Fort Meade Recovery: 89.6%

Riverview % Insol Received: 15.4%
New Wales % Insol Received: 16.0%
Bartow % Insol Received: 16.5%

Wingate Trucking Tons to Riverview: 871
Wingate Trucking Tons to New Wales: 0
Wingate Trucking Tons to Bartow: 0
```

Figure 2. Reported result for Mosaic's phospate production supply chain using data-driven optimization methods

they considered costs associated with processing and transportation, and used real data to develop correlations that were embedded within the optimization model (Figure 1 above). In the group's conclusions, they report "…this study has yielded valuable business insights and a flexible optimization model that can be utilized in the future for deeper investigations into the supply chain of phosphate production…These results can be fed into the larger discussion and investigation of Mosaic's cost model and provide the ability for the mine planning department to make data driven decisions."

This project highlights how the project-driven curriculum connects the two core courses of the data science certificate program and drives engagement with students and industry. The highlighted example project also shows that there is continuity and cohesiveness of topics built in the two core courses, which allows students to learn about how to process and analyze their data, and also to learn how to use the techniques from ChBE 4745/6745 to make optimal decisions or solve inverse problems. In other words, our two core courses are designed to "close-the-loop" between data collection, processing, analysis and decision-making.

Most students enjoyed the project, and some representative positive quotes are provided below:

- *The project was also really great and helpful in applying the concepts we were learning to a different practical setting. It helped me understand how I could use this more for work, etc. Would recommend keeping the project as is, the timelines kept well with the course.*

- *I thought the group project was a very well-designed component of the course.*

- *The course project was important for me to understand the course material in more detail.*

Vertically Integrated Groups:

Another innovative aspect of ChBE 4745/6745 and ChBE 4746/6746 is the use of vertically integrated project groups. The groups consist of undergraduate, graduate, and industry students. **This structure provides undergraduate students a chance to gain first-hand insight into graduate school and industry careers and provides industry students with the opportunity to mentor undergrads and learn from grad students**. The grad and industry students are

responsible for working together to provide the dataset and define the goals for the course project, and all group members are expected to delegate tasks and work together on completing the goals. Importantly, teams are required to meet for at least one hour every week and are encouraged to work together on homework and other assignments, and to complete a bi-weekly group evaluation to ensure that all members are participating and behaving professionally. **These requirements help drive engagement by requiring regular, synchronous interactions between students, and helps create a cohesive and supportive team where group members can learn from each other.** In ChBE 4746/6746, group pages were created on Canvas, and this provided a way for the instructors to communicate with each group and monitor activity. Overall, students had positive feedback regarding the group structure, as evidenced by the quotes below:

- *I liked the balance between graduate students/undergraduate students. Weekly meetings were very helpful with discussing homework and creating an open space for conversation. I'm glad the project was a semester long-process since projects for other classes tend to pile up at the end.*

- *I liked the way we worked in our group project. Weekly meeting was helpful to make the proper plan ahead of time and distribute/discuss about the responsibility among us.*

- *I also really liked that there were study groups, because I enjoyed getting to know some of my classmates and felt like they were a great resource.*

- *Your class was one of the few courses I enjoyed thoroughly. It was mainly due to the flexibility of the course and the topics covered in the course. I honestly wished that your course should have lasted a little longer. Though it was online, you made sure that by we interacted with each other by forming groups for projects and peer reviews.*

D-3: Course Content: Lectures, Jupyter Notebooks, Vocareum, and Peer Grading

Lectures: In both courses, the instructors worked with **the GT Professional Education office to record curated, high-quality short videos t**hat were released in a sequence of topics, or modules, every week. This structure was proven to be critical towards the success of the courses. Especially in the case of undergraduate and graduate students who are most likely taking asynchronous online courses for the first time, they often have initial doubts regarding the effectiveness of this delivery method. However, in all cases students found the collection of short theme-based recordings, coupled with weekly knowledge checks and all other assessments engaging and efficient.

Jupyter Notebooks: In addition to the recorded lectures, the course material in ChBE 4745/6745 is heavily based on **Python programming**, and **Jupyter Notebooks** are used in the course to present students with an integrated document that contains explanations of concepts along with functioning code. All lecture notes are provided in a set of Jupyter notebooks that students can access via Github, and are made permanently and publicly available so that students can access them even after the course is over. The course material in ChBE 4746/6746 is 30% lecture-based

(PDF and PowerPoint), and 70% Python programming notes, released to students in the form of Jupyter notebooks, and also made available via Github.

Vocareum: In addition, both courses make use of the Vocareum platform for hosting assignments and exams in Jupyter notebooks. Weekly homeworks include an auto-graded "skill check" where students complete coding exercises that can be submitted repeatedly for instant feedback, allowing them to keep trying tasks until they master the skills needed to successfully write and utilize coding tools. There is also a weekly homework assignment that asks more open-ended questions where students must apply their skills to problems similar to what they might encounter in their projects. **These assignments are peer graded**, allowing students to see how others solve the problem and get multiple perspectives on how to think about a dataset or apply different tools. Exams are also based on Jupyter notebooks hosted on Vocareum, and utilize a combination of auto-graded questions and instructor-graded open-ended questions, ensuring that students feel that the exams are consistent in style and content with their prior assignments. Some positive comments from students on these aspects are provided below:

- *Overall, the course felt streamlined. I think the exams worked out very well and they were fairly developed and also were more pragmatic rather than conceptual/derivations. I think this was a well rounded intro class and all aspects of material were useful (notes, videos, and homeworks). The focus on pragmatism was key in my opinion, especially with the inclusion of the outside folks.*

- *The course content is extremely well organized, and the course videos are high-quality, perfectly complimented by the detailed Jupyter Notebooks. I think this is an excellent example of a successful online course. I learned a lot from this class, and will definitely refer back to the class notes in the future.*

- *I think even though there was a lot of work required (skill check and homework), I think it was needed. The skill check helped understand how well you were learning from the lessons by having the assertions for auto grading and then the homework tackled a different section. So a lot of work but a really good way to understand the concepts. Grading other students homework was really helpful in seeing different ways classmates thought of how to solve the problem. Lots of work but really beneficial for understanding concepts and succeeding in the course!*

- *The lectures were easy to understand and well organized… the short module videos which never felt too much or over burdening.*

- *Although the content of this course was very difficult, Dr. Boukouvala did an excellent job of making sure that assignments, projects, and tests were fair and tested the knowledge we learned in the course (rather than previous knowledge some students had). I really enjoyed this course. At times I found it very challenging, but I feel that I learned a lot from it.*

- *The weekly knowledge checks were good. The homework was also reasonable and prepared us well for the exams.*

D-4: Electives

In addition to the two core cores, students in the certificate program take two elective courses, to expand their breadth and depth in their chosen areas of interest. The external students take online courses, especially those offered in the online MS in analytics, while the internal graduate students take in-person courses on the Atlanta campus. The following list of electives is provided to the students, although they can take additional courses with pre-approval from Professor Grover:

CS 6035	Introduction to Information Security
CS 6040	Computing for Data Analysis
CS 6220	Big Data Systems and Analytics
CS 6400	Database Systems Concepts and Design
CS 6421	Temporal, Spatial and Active Databases
CS 6430	Parallel and Distributed Database Systems and Applications
CS 6601	Artificial Intelligence
CS 7641/CSE 6740/ISYE 6740	Machine Learning
CSE 6140	Computational Science and Engineering Algorithms
CSE 6242	Data and Visual Analytics
ISYE 6402	Time Series Analysis
ISYE 6413	Design and Analysis of Experiments
ISYE 6414	Statistical Modeling and Regression Analysis
ISYE 6416	Computational Statistics
ISYE 6420	Introduction to Theory and Practice of Bayesian Statistics
ISYE 6501	Introduction to Analytics Modeling
ISYE 6644	Simulation
ISYE 6662	Discrete Optimization
ISYE 6664	Stochastic Optimization
ISYE 6740	Computational Data Analysis
ISYE 6810	Systems Monitoring and Prognostics
ISYE 7201	Production and Service Systems Engineering
ISYE 7203	Logistics Systems Engineering
ISYE 7406	Data Mining and Statistical Learning
MGT 6203	Data Analytics in Business
MGT 8823	Data Analysis for Continuous Improvement
PUBP 6501	Information Policy and Management
PUBP 6725	Information Security Policies and Strategies

> II. **Description of how the innovation has been evaluated**, including documentation of how the objectives were met, the extent to which the learning outcomes were achieved, the benefits derived from the innovation, and how they were measured.

The certificate program is assessed in accordance with policies established by Georgia Tech's the Office of Assessment.

Student Learning Outcomes and Metrics

<u>Objective 1:</u> Competence in Fundamentals of Data Science in Chemical Manufacturing
Students will develop and demonstrate competence in fundamental knowledge of data science relevant to chemical manufacturing.

<u>Objective 2:</u> Competence in Software Applications of Data Science in Chemical Manufacturing
Students will develop and demonstrate competence in using programming and software that uses data science to solve real-world problems in chemical manufacturing.

Final exams of core courses

In the first year of the program, to evaluate Objective 1, the score on the final exam of ChBE 6745 was analyzed to determine the number of students earning 80% or higher, with a target of 75% of students meeting this criterion. For the internal on-campus students, 88% (n above 80% = 15, total n = 17) of the learners received a final exam score of 80% or higher in ChBE 6745. For the external online students, 86% (n above 80% = 6, total n = 7) of the learners received a final exam score of 80% or higher in ChBE 6745.

For Objective 2, the score on the final exam of ChBE 6746 was similarly analyzed to determine the number of students earning 80% or higher, again with a target of 75% of students meeting this criterion. Among on-campus students, 76% (n above 80% = 16, total n = 21) of the learners received a final exam score of 80% or higher in ChBE 6746. For the online students, 86% (n above 80% = 6, total n = 7) received a final exam score of 80% or higher in ChBE 6746.

Thus, in the first year of the program, both the internal and the external cohorts performed above the target level in both objectives

Mid-program survey

A survey was conducted for external students in the first cohort of the program. Six of the seven students responded to the survey.

Who is enrolling in the program? When asked "What was your primary motivation for obtaining this graduate certificate?" four students indicated that it was to **expand their knowledge base**. One selected "change jobs/get a new job" and one selected "other." When asked "What is your current employment status?" five students indicated that they are **working full-time**, while one is a student.
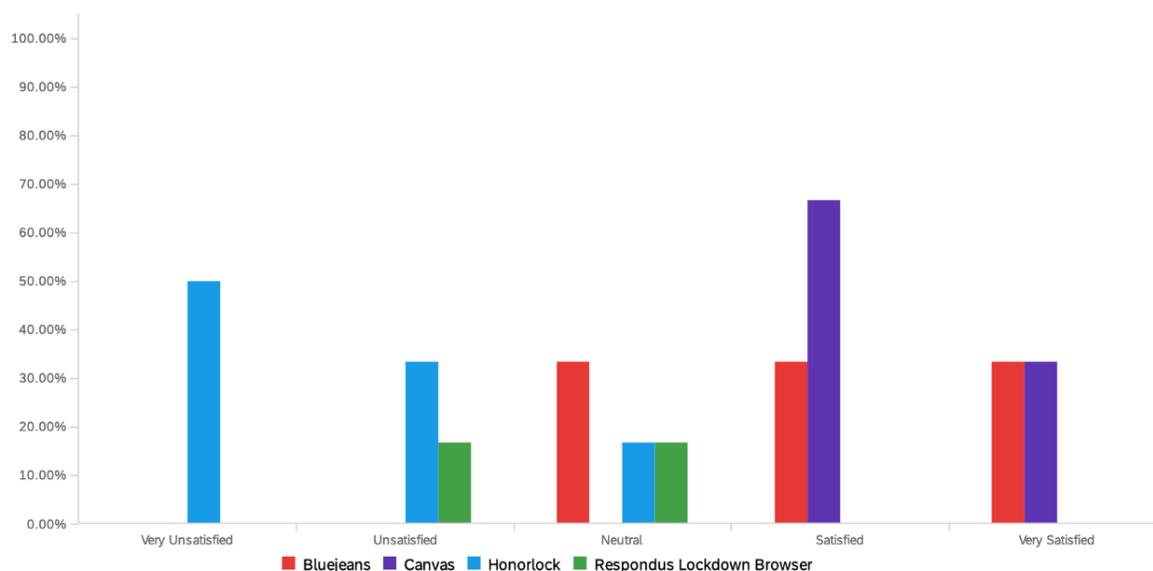
Course satisfaction:

Students were asked: "So far, how would you rate your satisfaction with the following aspects of course delivery?" The results are shown in Figure 3. Overall, students are most satisfied with the online lectures, the quality of feedback from instructors, and the number of opportunities to interact with course instructors. Since these are online courses, it is particularly notable that the students are satisfied with their interactions with the instructors. The students reported being less satisfied with the number of opportunities to interact with fellow students, which is being considered in the current group models of ChBE 6746 in this spring semester. Based on this feedback, the instructors are encouraging the students to meet with their project groups more often, and to utilize MS Teams to work with their group and the instructor.
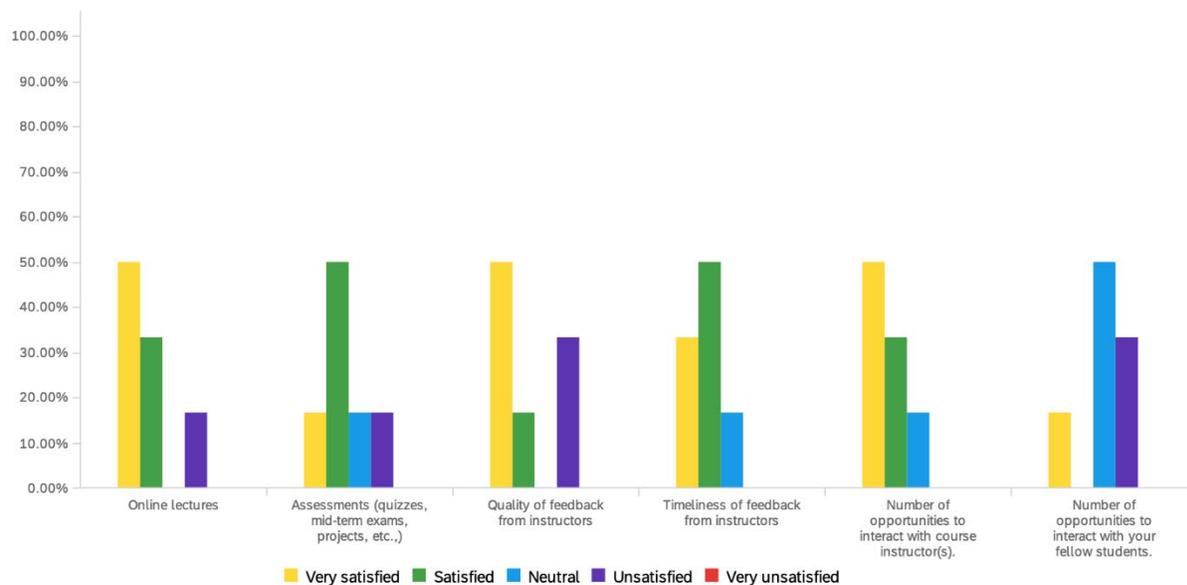
Figure 3: Response to "So far, how would you rate your satisfaction with the following aspects of course delivery?"

Another question dealt with the instructional technology, as shown in Figure 4. Students were most satisfied with Canvas and least satisfied with Honorlock. In subsequent semesters, the instructors reduced or even eliminated the used of proctoring software, and incorporated honor code statements to exams.

Figure 4: Responses to "Please rate your satisfaction with the following course technologies."

Students were also asked "So far, how would you rate your experience in the Online Graduate Certificate in Data Science for the Chemical Industry?" Three students indicate that the program exceeds or far exceeds their expectations, two that it meets expectations, and one that it falls short of expectations. Based on this survey together with direct student feedback, the program is continuing to build upon its successful early performance in order to continue to serve student needs.

III.   Potential for others to adopt or adapt

As one of the first stand-alone graduate certificates at Georgia Tech, this program is already proving to be a model for others to adopt, partnering with professional education for this online asynchronous format.

The Graduate Certificate in Data Science for the Chemical Industry is Georgia Tech's first fully online graduate certificate. By benefiting both off campus and on campus students, it creates a template for much broader innovation in graduate education at GT that is likely to appeal to a very large group of students for whom a complete online MS is too large of a commitment.

Nomination for the CTL 2022 Curriculum Innovation Award:
DSCI Team (ChBE Faculty Drs. Boukouvala, Grover, Medford, Meredith, and Sholl)

Jennifer Wooley, Director of Academic Programs and Services at Georgia Tech's Department of Professional Education, agrees that this graduate certificate program is a groundbreaking offering at Georgia Tech that is already inspiring others to adopt its core concepts and methods:

> *"As leaders in various areas of engineering wrap their heads around the possibilities big data offers, and how deeply it can impact their viability in the marketplace, they have a similar sense of urgency.*
>
> *Although DSCI launched in Fall 2020 and is the first online graduate certificate offered by Georgia Tech, the Institute's leadership already sees the benefit of providing this powerful combination of customized knowledge and skills. **In fact, DSCI's immediate success convinced them to approve two more graduate certificates that use this pairing concept**. Those programs will launch later this year. Additionally, a team is exploring and validating other variations for future rollouts." – Jennifer Wooley, Professional Education at GT*

External audiences have also expressed great interest in this program. For instance, Dr. Medford gave a talk about this program at the AIChE Annual Conference in 2021, in a session on "Teaching Data Science to Students and Teachers." The slides are archived here:
https://cache.org/teaching-data-science-aiche-2021
It was the only presentation that focused on using an online asynchronous format to teach data science to working professionals. The team was invited to write a paper on the program to the special issue of *Chemical Engineering Education*. (submission planned in February)

Additionally, Dr. Boukouvala was invited to the Computing Aids in Chemical Engineering (CACHE) workshop for the Future of Cyber-Assisted Education as a rising star young educator, where she presented an overview of the certificate and her course. Based on this talk, she co-authored an article in *Chemical Engineering Education* on teaching courses using Python Jupyter Notebooks.

The positive reception to these presentations, as well as published and upcoming articles detailing the program and its impact, speaks to a great opportunity to adopt this program both elsewhere at Georgia Tech and at other institutions. Having a strong partnership with Professional Education, as we do here at GT, is a key required resource if other schools or universities wish to develop a similar program.

References

[1]   Ayers, Ryan "Four things to know about big data in chemical engineering", **2017**, https://www.aiche.org/chenected/2017/08/four-things-know-about-big-data-chemical-engineering
[2]   Qin, S.J., Chiang, L.H. "Advances and opportunities in machine learning for process data analytics" *Computers & Chemical Engineering* **2019** *126*, 465-473.

Cristina U. Thomas, Ph.D.
Senior Laboratory Manager
Global Research & Development Services
Research & Development Global Process Owner

3M Center Building 0220-07-E-24
St Paul, MN 55144-1000
651-736-2185 Office
cuthomas@mmm.com

3M

February 1, 2022

Dear CTL Awards Committee,

We are writing in support for the nomination of Drs. Martha Grover, Carson Meredith, David Sholl, Fani Boukouvala, and AJ Medford for the 2022 GT CTL Curriculum Innovation Award. As the exponential speed for digitization emerges in industry, professionals trained in chemistry, chemical engineering, and/or working in the chemical industries, have felt the need to rapidly update their skills in Data Science and Data management. There was clearly a gap of certificates and master's degrees for professionals that, while building in the fundamentals of data science, also had practical examples pertinent and relevant to the chemical enterprise. Throughout many interactions between 3M and Georgia Tech staff, we learned that the Chemical Engineering Department was working on adding an online certificate in Data Science for the Chemical Industries. We immediately communicated our excitement and our desire to participate as we were already looking for online and practical programs to upskill our researchers and developers and started working with Prof David Sholl and his team. It will be good to highlight a quote from two of our students regarding reasons for joining the certificate program:

"I joined the program since I'm involved with developing products with integrated sensors. I wanted to make sure that I was taking steps to being able to handle the data from these products."

"Wanted to join the program since I work at a manufacturing plant and work with a lot of data. We have dataloggers for all products that track every aspect of the product and wanted to learn/improve on ways to analyze that data and trends. "

 Our support for these students has yielded the anticipated results. These students are innovators and are asked to incorporate digital technologies in their research. We can attest to the impact to innovation of the courses in the certificate. It has been an engaging and challenging learning experience requiring students to go back and review prior courses and learn new programing languages. It is clear to us the learning is real and impactful and is yielding results in equipping our researchers with state of the at digital practices. It is best to hear from them directly to clearly illustrate the impact in innovation.

"I found Georgia Tech's Data Science certificate program very useful. I work in R&D and I had heard a lot of buzz around data science and how it's going to change the way we do research. The program was very informative and put everything in context. I learned a lot from the program. Thank you to the instructors for designing a program targeted towards chemical engineers. We are already using some of these tools to get more insights in our extrusion processes."

"The program is well thought out to help domain experts get the tools needed for deeper data understanding. It is structured in a way that is flexible enough for professionals, while keeping the content challenging and engaging. I have no doubt this will help open up many options for future career development."

We would like to emphasize once again the impact of the quality of teaching and learning at Georgia Tech as 3Mers have taken classes offered by this certificate. We gathered very positive feedback about the quality of the courses and instruction:

"The coursework has been instrumental in my computer programming and data analysis. I feel like I've become proficient in programming real-world data analyses in Python, which is a leap from where I was beforehand."

"This program helps me lead the AI/modeling project team at 3M to leverage data science in chemistry accelerating our new product development.

"Extremely valuable. Perfect mix of background materials/coding and applications. Great use of real-world examples to teach."

We lead the 3M Global Technical Learning Academy and are responsible for offering valuable and practical learning experiences. Our interest in the program resulted in 3M funded scholarships for 3M participants with the condition that they secured acceptance to the Georgia Tech Program. We value this program so highly that 3M Corporate R&D has granted full 'internal' scholarships to eleven employees to date and plan to award more for the coming school year with full support from their managers. Recipients of the 3M scholarship indicate that:

"The management team expect leveraging the data science techniques to accelerate new product development."

"Management knows that for 3M, Science applies to life, and in order to support your qualification work, data needs to be used to help with your selections. Data is heavily relied upon, and management has been really great in taking the time to improve our current skillsets."

We want to reiterate our full support to the nomination of the founders of the Data Science in the Chemical Industry certificate and believe that Drs. Grover, Meredith, Sholl, Boukouvala, and Medford are well deserving of such prestigious recognition due to the impact to rapid digitization within the chemical industries.

Sincerely,

Cristina U Thomas
Global R&D Services and R&D Global Process Owner
3M Corporate Research and Development

Chris Jacobs
Global Technical Learning Academy Leader

Aaron Liu
School of Chemical & Biomolecular Engineering
311 Ferst Drive, N.W.
Atlanta, Georgia 30332-0100 U.S.A.
aliu319@gatech.edu

January 28, 2022

Dear CTL Awards Committee,

It is with great pleasure that I am supporting *Drs. Martha Grover, Carson Meredith, David Sholl, Fani Boukouvala, and AJ Medford* in their nomination for the 2022 GT CTL Curriculum Innovation Award. I enthusiastically believe they deserve recognition for designing and mobilizing the Data Science for the Chemical Industry (DSCI) certificate program within the School of Chemical & Biomolecular Engineering (ChBE). As a Ph.D. candidate and graduate student in ChBE, I am participating in the DSCI program to fulfill a data science minor and pursue a data-driven role for a chemical or materials company.

From my perspective as a current student and former engineering employee for the chemical industry, the program is timely because it addresses both a burgeoning demand for domain-specific data analytics expertise by the chemical industry as well as booming interest in data science from students and academia. **Additionally, the curriculum is effective because it is innovative, well-curated, and engaging**. The interdisciplinary curriculum bridges expertise between data science and chemical engineering in a way that offers unique, tailored knowledge that the two disciplines cannot provide independently of one another. The core curriculum takes chemical engineers with originally little coding expertise and transforms them into chemically fluent data scientists with the skills to revolutionize smart manufacturing processes in energy, advanced materials, and other chemical processes essential to modern society. Although this integrated skillset is valuable and in high demand at the industry level, some of my former classmates could not find a similar enriching curriculum when searching at other universities.

To exemplify the curriculum's engagement, all three of the courses I have completed in the program thus far have incorporated a course project that has challenged me to apply data analytics skills to highly relevant chemical datasets of my choice. For example, one of the datasets was provided by Dow Chemical, adding an exciting real-world element to the challenges we explored. **Some of the work and skills from these projects has even directly contributed to the publication of two peer-reviewed papers** that I had the pleasure of co-authoring in the past couple of years. Furthermore, I had an opportunity in 2021 to apply my newfound data analytics skillset at a summer internship at BASF. Armed with new perspectives from the curriculum, I drove my project to completion by leveraging data-driven methods to construct a smart data modeling interface for fuel and lubricant products, which would not have been possible without DSCI.

Finally, I believe that the DSCI program is broadly impactful because it paves the way for interdisciplinary professional programs within other schools and fields within Georgia Tech. Data science and machine learning are clearly becoming a dominant paradigm within technologies outside of chemical engineering; therefore, some of the successful engagement methods within DSCI are a fruitful opportunity to excite students outside of ChBE about applications within their field. These factors and more qualify the professors listed above and the DSCI for selection to the 2022 GT CTL Curriculum Award. Please feel free to contact me at aliu319@gatech.edu if I can provide any further information.

Sincerely,

Aaron Liu
Ph.D. Candidate, School of Chemical & Biomolecular Engineering

19Jan2021

Dear CTL Awards Committee,

My name is Zak Kuiper and I have just completed the Industry Certificate for Data Science in the Chemical Industry as of the 2021 Fall semester. I am also a Principal Data Scientist with The Mosaic Company, a F500 chemical company. I am writing today in support of the nomination of the faculty who developed and executed this fantastic program I was fortunate enough to be a part of.

When this program was first introduced to me a few years ago I had just started on my path as a Data Scientist helping my organization stand-up its first internal team dedicated to Data Science & Machine Learning development. It turned out to be a major impactor in not only my own professional development but indeed the development of the company's strategy around how we would approach and carry our new analytics group. So much so that on a weekly basis I would bring notes from my studies to our staff meetings to help build the group's base understanding as we developed. This program very quickly began illuminating how different dealing with data and use cases from the chemical industry is from the vast majority of the lessons and use cases taught as foundational to those looking to learn Data Science. Our industry is one who can greatly benefit from the implementation of these skills however it is so different and un-represented in most of the material available that it can be quite daunting for Chemical Engineers to begin learning and practicing. A program like this is truly filling a need in a way that I feel most of the industry didn't even consider was possible.

One of the biggest things this program has done exceptionally well is just how practical it is. This is undeniably illustrated by the fact that class projects I worked on in Dr. Medford's and Dr. Boukouvala's courses have become tangible funded projects in my organization which show up on our groups value capture reports. These two courses were so well done, I took my class project from Dr. Medford's course, and expanded on it based on the learnings directly from Dr. Boukouvala's. Resulting in my company hiring a 3rd party team to further develop and scale this project across the business and we expect it to return around $8M in EBIDA uplift beginning this year.

In closing I would just like to re-iterate my point that this program is first to the table in answering a need which most of the chemical industry is just realizing they have. It has allowed me to develop as a professional significantly quicker than I would have otherwise and has shaped how our organization has focused on the area of Data Science. Best of all it's targeted practical instruction style has directly led to immediate real-world returns making it what any true industry professional cares about most, a fantastic return on investment.

Thank you for your time,

Zak Kuiper

Data Scientist; NA Digital Strategy

The Mosaic Company

Student Letter of Support

Jackson Dean
Georgia Tech B.S. ChBE 2021
Data Science Associate – Pharmaceutical Supply Chain
GlaxoSmithKline (GSK)

Dear CTL Awards Committee:

I am delighted to write this letter of support for the nomination of the team of faculty behind the Data Science for the Chemical Industry certificate program for the Curriculum Innovation Award. I was a student in Dr. Medford's "Data Analytics for Chemical Engineers" in Fall 2020 and Dr. Boukouvala's "Data-Driven Process Engineering" in Spring 2021. Both courses gave me a solid foundation in Python programming, machine learning, and applications of data science in the chemical industry. I would consider these courses the most important classes I took in undergrad because they helped me land my first job in data science and inspired me to further study data science via Georgia Tech's Online Master of Science in Computer Science (OMSCS), which I am currently enrolled in.

Currently, I work in pharmaceutical manufacturing and use data-informed techniques to improve our processes. This industry, as well as the broader chemical industry, is experiencing rapid change towards automation, where data plays a central role. While the core ChBE curriculum teaches concepts such as first principles and process design, a fresh graduate entering the industry will quickly need to become "data-fluent" to carry out their duties- often statistical process monitoring or communicating the success of a project through data visualization. The curriculum of this certificate program directly addresses this need of creating "data-fluent" chemical engineers- teaching skills such as data management/cleaning, building machine learning models, and data visualization. Additionally, jobs for entry-level process engineers often list Python programming as a qualification, which is the coding language of choice used for all assignments and projects in these courses but is not taught in a traditional chemical engineering curriculum.

I found success in these courses due to the thoughtful design of each course, which fostered a high level of engagement despite it being an online course. The combination of lecture videos and instructor-created code demo notebooks to act as the primary course material meant that I had multiple ways of learning the material. I used the videos to learn the concepts verbally from the instructor, and read through the code to understand how the concepts are implemented in Python. The instructors were very present to help diagnose any problems related to "setting up" Python on their computers, which can be difficult for someone new to the data world. The most engaging parts of these courses were the homework assignments, which required us to write our own code - arguably the most valuable skill in this course. The data sources used in these assignments were always somehow related to chemical engineering (infrared spectra, sensor data from a Dow distillation process, reactor optimization, etc.). Not only are programming and statistical concepts put into practice, they are applied to data a chemical engineer is likely to encounter.

The courses also involved group projects that allowed students to identify their own project statements, research data sources, and use techniques learned in the class to develop data-driven insights. I especially appreciated the format of these projects: the instructors set iterative deadlines to ensure we were dedicating appropriate efforts to each part of the analytics process. For example, we had to identify our data source and make sure it is usable early in the semester. We then had to develop early-stage analytics models and report on their performance, then get feedback on how to improve the models well before the end of the project. This framework allows students to understand the entire process of building an analytics pipeline, getting iterative feedback, and continuously improve their project. This is very similar to how analytics projects are done in the industry.

Overall, these courses teach skills that traditional chemical engineers often lack and can help students land data-based rolls in the chemical industry and beyond. The high-quality design of the courses and the instructors' effort to engage with the students in an online format make the faculty excellent candidates for this award.

Sincerely,

T. Jackson Dean III